# Bioprocess development for terpenoid production via the methylerythritol phosphate pathway in *Escherichia coli*

## Dissertation

A dissertation accepted by the Faculty of Energy-, Process- and Bio-Engineering of the University of Stuttgart in partial fulfillment of the requirements for the degree of Doctor of Engineering Sciences (Dr.-Ing.)


By:

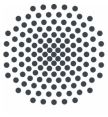
**Vikas Patil**

from Nagpur, India


Supervisor: Prof. Dr.-Ing. Ralf Takors

Co-examiner: Prof. Dr. Georg Sprenger

Date of oral examination: 23.11.2021




University of Stuttgart
Germany

Institute of Biochemical Engineering


2022

## Acknowledgments

This research was conducted at the Institute of Biochemical Engineering (IBVT) at the University of Stuttgart and began in July 2016. Many people have contributed to the progress of my work and life at the IBVT. First, I would like to express my sincere gratitude to my advisor and mentor, Prof. Dr.-Ing. Ralf Takors, for the continuous support of my Ph.D. study and related research as well as his patience, motivation, and immense knowledge. His guidance helped me throughout the research and writing of this thesis. I could not have imagined having a better advisor and mentor for my Ph.D. study. I am well and truly indebted to him for his mentorship. My sincere thanks also go to Dr. Georg Sprenger and Dr. Bernhard Hauer for their interest in the research and participation in the examination committee. I would also like to extend my gratitude to Dr. Martin Siemann-Herzberg, Dr. Björn Voß, and Dr. Bastian Blombach for their continued guidance with my work. I also want to thank Manus Bio Inc. in Cambridge, USA for their financial support and for giving me the opportunity to support their research endeavor at IBVT. Additionally, I would like to extend a special thanks to Dr. Ajikumar Parayil and Dr. Christine Santos for their guidance on various occasions during this project.

Further gratitude goes to Julia Zieringer, Max Becker, and Ryan Fahs for being a part of this project. Julia Zieringer was responsible for transcriptomics data analysis during the project, while Max Becker and Ryan Fahs supported the scale-up process to a 300L bioreactor. I would also want to thank my student, Massimo Villa, for contributing to this study.

Also, I thank all my present and past coworkers for providing a wonderful atmosphere at the IBVT, especially Maike Kuschel, Salaheddine Laghrami, Lisa Junghans, Andreas Ankenbauer, Maria Hermann, Andreas Freund, Alexander Dietrich, Natascha Verhagen, Martin Ziegler, Marius Braakmann, and Andy Wiranata Wijaya.

Last but not the least, I would like to thank my family—my parents, my brother Yuvraj, and my lovely wife Nupoor—for supporting me throughout my thesis and life in

general.

# Contents

# List of Figures

# List of Tables

# Nomenclature

| | |
|---|---|
| $\dot{n}_{O2}$ | Molar flux of oxygen, mol/L |
| $\dot{V}_{gas}$ | Volumetric flow of air/exhaust gas, L/h |
| $\mu$ | Specific growth rate, $h^{-1}$ |
| $\rho$ | Density of liquid, g/L |
| $c_F$ | Concentration of substrate in feed, g/L |
| $c_{p,t}$ | Concentration of product in bioreactor at time t, g/L or %AU/L |
| $c_p$ | Concentration of product in bioreactor, g/L or %AU/L |
| $c_{s,t}$ | Concentration of substrate in bioreactor at time t, g/L |
| $c_s$ | Concentration of substrate in bioreactor, g/L |
| $c_{x,t}$ | Concentration of biomass at time t, g/L |
| $c_x$ | Concentration of biomass, g/L |
| $di$ | Impeller diameter, m |
| $F_{in}$ | Feed rate of substrate, L/h |
| $k_La$ | Volumetric mass transfer coefficient, $h^{-1}$ |
| $m_p$ | Mass of product, g or %AU |
| $m_s$ | Mass of substrate, g |

$m_x$ Total biomass, g

$n$ Impeller agitation speed, RPM

$N_p$ Power number of impeller

$P$ Power input in the bioreactor, Watt/L

$p$ Pressure of the mass flow controller, atm

$p_{O2}$ Dissolved oxygen level, % of air saturation

$q_{AA}$ Biomass specific consumption rate of amino acid, g/g/h

$q_{AMD}$ Specific productivity of AMD4,11, mg/g/h

$q_{EGN}$ Specific productivity of EGN, %mAU/g/h

$q_{glucose}$ Biomass specific consumption rate of glucose, g/g/h

$q_{O2}$ Biomass specific consumption rate of oxygen, mol/g/h

$q_p$ Specific productivity of product, g/g/h or %mAU/g/h

$q_s$ Specific consumption rate of substrate, g/g/h

$T$ Temperature of mass flow controller, °C

$v_g$ Superficial gas velocity, m/h

$V_{r,t}$ Reaction/recipe volume at time t, L

$V_r$ Reaction/recipe volume, L

$y_{CO2}$ Volumetric fraction of carbon dioxide in air/exhaust gas

$y_{O2}$ Volumetric fraction of oxygen in air/exhaust gas

$Y_{sx}$ Composition of substrate in biomass, g/g

$Y_{xs}$ Yield of biomass on substrate, g/g

AMD4,11 Amorpha-4,11-diene

CER Carbon evolution rate, mol/L/h

DMAPP Dimethylallyl diphosphate

DOX 1-Deoxy-d-xylulose

EGN Terpenoid-name not to be disclosed

FBA Flux balance analysis

G3-P/GAP Glyceraldehyde 3-phosphate

IPP Isopentenyl diphosphate

MEcPP 2-C-Methyl-D-erythritol-2,4-cyclopyrophosphate

NADH Nicotinamide adenine dinucleotide

NADPH Nicotinamide adenine dinucleotide phosphate

OD600 Optical density

OTR Oxygen transfer rate, mol/L/h

OUR Oxygen uptake rate, mol/L/h

RQ Respiratory quotient

vvm volumetric flow rate of air per unit volume of media in the bioreactor

YE Yeast extract

DCW Dry cell weight, g

## Zusammenfassung

Terpenoide sind eine Familie natürlich vorkommender Verbindungen, von denen bis heute über 80.000 identifiziert wurden. Während die meisten Terpenoide in Pflanzen vorkommen, können sie ebenfalls in Insekten, Bakterien und Pilzen auftreten. Sie stammen aus der Verknüpfung zweier Bausteine: Isopentenyldiphosphat (IPP) und dessen Isomer Dimethylallyldiphosphat (DMAPP). Beide können über die Wege Methylerythritolphosphat (MEP) und Mevalonat (MVA) erzeugt werden. Der MVA-Weg ist in Eukaryoten weit verbreitet und wurde erstmals in den 1950er Jahren entdeckt. Daher stand er bis zur Entdeckung des MEP-Wegs im Mittelpunkt der Forschung zur Terpenoidproduktion. Der MEP-Weg wurde in den 1990er Jahren entdeckt und ist auf Kohlenstoff bezogen theoretisch effizienter, mit einer um 20% höheren molaren IPP-Ausbeute für Glukose als Kohlenstoffquelle. Aufgrund komplexer Regulationsvorgänge und der Energieineffizienz sind die Forschungsbemühungen zur Erfassung des vollen Potenzials jedoch noch nicht voll ausgeschöpft worde n. Der MEP-Weg ist redoxausgeglichen und erzeugt daher kein überschüssiges Nicotinamidadenindinukleotid (NADH) für die Herstellung von Adenosintriphosphat (ATP), das für die Zwischenschritte des Wegs erforderlich ist. Dies ist zum einen mit Herausforderungen im Zusammenhang mit der Bereitstellung von Vorläufern verbunden, zum anderen mit Engpässen, die mit der begrenzten Aktivität von Zwischenenzymen wie IspG und IspH zusammenhängen. In dieser Studie wurden Probleme im Zusammenhang mit der Herstellung von zwei Terpenoid- / Terpenprodukten angegangen: EGN (eine Mischung aus Valencen, Nootkatol und Nootkaton) und Amorpha-4,11-dien. Unser *Escherichia coli*-Stamm wurde entwickelt, um die MEP-Weg-Enzyme Dxs, Dxr, IspF, IspG und IspH zu überexprimieren und hierbei den Kohlenstoffffluss aus Pyruvat zu ziehen. Der EGN1-Stamm wurde mit der konstitutiven Expression von Valencen-Synthase und

-Oxidase konstruiert, um EGN zu produzieren. Der AMD-stamm wurde als Gegenstück zur Amorpha-4,11-dien-Produktion entwickelt. Ein Ungleichgewicht in der Expression von *ispf* und *ispg* wurde als Ursache für den Einbruch des Prozesses mit EGN1 identifiziert. Dieses Ungleichgewicht bewirkte einen Abfluss von 2-C-methyl-D-erythritol-2,4-cyclopyrophosphat (MEcPP), einem Zwischenprodukt des MEP-Wegs, welcher in einen erhöhten Bedarf an Phosphat resultierte. Phosphatmangel wiederum führt zur Laktatbildung, einer erhöhten spezifischen Glukoseaufnahmerate und übermäßigem Schäumen im Bioreaktor. Eine Erhöhung der Phosphatzugabe um 50%, gesteuert über die Glukosezufuhr, erzeugte eine Verbesserung des EGN-titers um 65% (165,22±3,72 %AU/L). Die Laktatbildung wurde zudem um 100% reduziert. Dieser Prozess wurde daraufhin in einen 300 L Bioreaktor mit einem Titer von 177,03 %AU/L hochskaliert. Die Sauerstofflimitierung machte 51,5% der Abnahme des EGN-Titers mit 8,99 g/L Nebenproduktbildung in Form von Acetat aus. Danach verlagerte sich der Schwerpunkt der Arbeit auf die Amorpha-4,11-dien-Produktion, gefolgt von einer anschließenden Neubetrachtung von EGN. Bei der Betrachtung von Nicotinamidadenindinukleotidphosphat (NADPH) und Pyruvat, die mehrere Reaktionen des MEP-weges unterstützen, wurde festgestellt, dass die Einstellung eines geeigneten Verhältnisses von Glukose- zu Sauerstoffversorgung notwendig ist, um die Verfügbarkeit von NADPH und Pyruvat für den MEP-Weg zu verbessern. Dies wurde unter Verwendung des AMD-Stammes demonstriert, der in zwei Systemen untersucht wurde: einem einphasigen (wässrigen) und einem zweiphasigen Systeme (unter Zusatz von Öl). Die Einstellung von $q_{O2}/q_{Glucose}$ führte zu einer 6,7-fachen Verbesserung unseres Titers. Unser derzeitiges Verfahren erzeugt Amorpha-4,11-dien mit 2,44 mg/g/h und einer C-mol-Ausbeute von 5,08% auf Glukose. Transkriptomdaten stützten die Beobachtung, dass eine höhere Verfügbarkeit von Glukose zu einer Herunterregulierung des Aminosäurebiosynthesewegs führte. Wie bei der Flussbilanzanalyse festgestellt, zeigte das zweiphasige System, dass die Hochregulierung der Transhydrogenase *pntAB* zu einer NADPH-Limitierung führt. Dies erklärte weiter die Akkumulation von DOX im zweiphasigen System mit einer höheren spezifischen Glukoseaufnahmerate. Die

Akkumulation des ATP-Pools im System mit einem höheren Verhältnis $q_{O2}/q_{Glucose}$ stimmte mit der hochregulierten Expression von Genen überein, die an der oxidativen Phosphorylierung beteiligt sind. Der optimale Bereich wurde, unter Verwendung der Flussbilanzanalyse über die COBRA-Toolbox, bei einer spezifischen Glukoseaufnahmerate von 3–5 mmol/g/h identifiziert. Eine maximale Produktivität von 0,532 mmol/g/h wurde bei einer spezifischen Glukoseaufnahmerate von 3 mmol/g/h und einer spezifischen Sauerstoffaufnahmerate von 3,71 mmol/g/h beobachtet. Um den entwickelten Basisprozess weiter zu verbessern, untersuchten wir, ob ein komplexes Substrat (Hefeextrakt) die Verfügbarkeit von Pyruvat verbessern und die mit NADPH verbundenen Kosten senken würde. Als Folge konnte der Amorpha-4,11-dien-Titer um 72% auf 5,24 g/L verbessert werden. Der Abbau von Aminosäuren führte zu einer Verdopplung der Produktivität. Aspartat, Glutamat, Serin, Glycin, Threonin, Arginin, Alanin und Prolin wurden als Aminosäuren identifiziert, die die Amorpha-4,11-dien-Produktion positiv beeinflussen. Da die Ergebnisse der Prozessentwicklung bisher speziell auf den MEP-Weg ausgerichtet waren, haben wir die Ergebnisse erfolgreich auf die EGN-Produktion für den EGN2-Stamm übertragen, der ähnlich wie AMD über einen anderen Downstream-Weg für die entsprechende Terpenoid-Produktion verfügt. Im EGN2- Stamm wurden die Gene *ispG* und *ispH* überexprimiert, was zu einer verringerten MEcPP-Akkumulation führte. EGN2 zeigte einen um 75% niedrigeren Phosphatverbrauchs aufgrund potenziell reduzierter MEcPP-Pools. Das für Amorpha-4,11-dien entwickelte Verfahren führte bei Verwendung des EGN2-Stammes zu 292,55±1,59 %AU/L EGN. Schließlich haben wir uns mit der Bedeutung der Sauerstoffübertragungsrate für ein erfolgreiches Scale-up der Terpenoidproduktion über den MEP-Weg befasst. Wir erhöhten den Leistungseintrag um das 3,7-fache und die Begasung um 50%, was in einer 3,27-fachen Erhöhung des Stoffübergangs von Sauerstoff resultierte. Indem wir den volumetrischen Leistungseintrag und die Gasleerrohrgeschwindigkeit als kritische Faktoren für den Stoffübergangskoeffizienten von Sauerstoff wählten, skalierten wir das entwickelte Verfahren auf ein Arbeitsvolumen von 300 L mit einem Produktionsverlust von lediglich 18%.

## Summary

Terpenoids are a family of naturally occurring compounds, with over 80,000 identified to date. While the majority of terpenoids are found in plants, they also occur in insects, bacteria, and fungi. They are derived from the condensation of two building blocks: isopentenyl diphosphate (IPP) and its isomer dimethylallyl diphosphate (DMAPP). Both can be generated by the methylerythritol phosphate (MEP) and mevalonate (MVA) pathways. The MVA pathway is prevalent in eukaryotes and was first discovered in the 1950s. Therefore, it was the primary focus of research on terpenoid production until the discovery of the MEP pathway. The MEP pathway was discovered in the 1990s and is theoretically more carbon efficient, with a 20% higher molar yield of IPP for glucose as a carbon source. However, due to the complexities of regulations and energy inefficiency, R&D efforts to register its full potential have been lagging. The MEP pathway is redox balanced and thus fails to generate excess nicotinamide adenine dinucleotide (NADH) for the production of adenosine triphosphate (ATP), which is required for the intermediate steps of the pathway. This comes with challenges related to the supply of precursors and bottlenecks associated with the limited activity of intermediate enzymes such as IspG and IspH. In this study, we attempted to sequester and address issues related to the production of two of our terpenoid/terpene products: EGN (a blend of valencene, nootkatol, and nootkatone) and amorpha-4,11-diene. Our *Escherichia coli* strain is engineered to overexpress the MEP pathway enzymes Dxs, Dxr, IspF, IspG, and IspH to pull carbon flux from pyruvate. The EGN1 strain was engineered with the constitutive expression of valencene synthase and oxidase to produce EGN. The AMD strain was engineered as its counterpart for amorpha-4,11-diene production. An imbalance of *ispF* and *ispG* expression was identified as the culprit for the breakdown of the process with EGN1.

This resulted in an efflux of MEP pathway intermediate 2-C-methyl-D-erythritol-2,4-cyclopyrophosphate (MEcPP), thereby resulting in an increased demand for phosphate. Phosphate starvation led to lactate formation, an increased specific glucose uptake rate, and excessive foaming. Increasing phosphate addition by 50% via glucose feeding led to a 65% improvement in EGN titer (165.22±3.72 %AU/L). Lactate formation was reduced by 100%. This process was then scaled up to a 300L bioreactor with a titer of 177.03 %AU/L. Oxygen limitation accounted for 51.5% of the depreciation of EGN titer with 8.99 g/L of byproduct acetate formation. Thereafter, the focus of the thesis shifted to amorpha-4,11-diene production followed by a subsequent revisitation to EGN. Upon identifying the significance of nicotinamide adenine dinucleotide phosphate (NADPH) and pyruvate supporting multiple reactions of the MEP pathway, it was established that installing an appropriate ratio of glucose to oxygen supply was necessary to enhance the availability of NADPH and pyruvate for the flux of the MEP pathway. This was showcased using the AMD strain, which was studied using two systems: monophasic (aqueous) and biphasic systems (with the addition of oil). The installation of apt $q_{O2}/q_{glucose}$ resulted in a 6.7-fold enhancement of our titer. Our current process produces amorpha-4,11-diene at 2.44 mg/g/h with a 5.08% C-mol yield on glucose. Transcriptomics data supported observations that identified downregulation of the amino acid biosynthesis pathway due to a higher availability of glucose. As observed with flux balance analysis, the biphasic system observed that the upregulation of transhydrogenase *pntAB* results in NADPH limitation. This further explained the accumulation of DOX in the biphasic system with a higher specific glucose uptake rate. The accumulation of ATP pools in the system with a higher ratio $q_{O2}/q_{glucose}$ was congruent with the upregulated expression of genes involved in oxidative phosphorylation. The window of optimum was identified at a specific glucose uptake rate of 3–5 mmol/g/h using flux balance analysis via the COBRA toolbox. A maximum productivity of 0.532 mmol/g/h was observed for a specific glucose uptake rate of 3 mmol/g/h and specific oxygen uptake rate of 3.71 mmol/g/h. To further improve upon the developed baseline process, we examined whether a complex sub-

strate (yeast extract) would improve pyruvate availability and reduce expenses related to NADPH. We improved our amorpha-4,11-diene titer by 72% to reach 5.24 g/L. The catabolism of amino acids resulted in a significant 2-fold productivity improvement. Aspartate, glutamate, serine, glycine, threonine, arginine, alanine, and proline were identified as amino acids that positively impact amorpha-4,11-diene production. Since the process development findings were particular to the MEP pathway until now, we successfully transferred the results to EGN production for the EGN2 strain, which is similar to AMD has a different downstream pathway producing the corresponding terpenoid. EGN2 overexpressed the *ispG* and *ispH* genes, resulting in reduced MEcPP accumulation. EGN2 resulted in a 75% reduction in phosphate usage due to potentially reduced MEcPP pools. The process developed for amorpha-4,11-diene resulted in 292.55±1.59 %AU/L of EGN when using the EGN2 strain. Finally, we addressed the importance of oxygen transfer rate to a successful scale-up of terpenoid production via the MEP pathway. We used a 3.7-fold increase in power input and increased aeration by 50%, which resulted in a 3.27-fold increase in the mass transfer of oxygen. By selecting volumetric power input and superficial gas velocity as the critical factors enhancing the mass transfer coefficient of oxygen, we scaled up the developed process to a 300L working volume with a mere 18% loss of production.

# Chapter 1

# Introduction

## 1.1 Biotechnology

Chemicals play an important role in every aspect of modern living. The chemical industry encompasses over 70,000 different products that contribute to the global economy. According to a report, the chemical industry touches nearly every good-producing sector and makes an estimated $5.7 trillion contribution to the world's gross domestic product (GDP) through direct, indirect, and induced impacts equivalent to 7 % of global GDP while supporting 120 million jobs worldwide (ICCA & Oxford Economics, 2019). As such, chemicals are ubiquitous in our lives. Despite playing an instrumental role in the industrial revolution, the chemical industry is currently facing major challenges related to the sustainability of its operations. The derivation of raw material from non-renewable fossil fuels is responsible for the fickle nature of raw material pricing. Additionally, most processes involve numerous major processing steps and require large facilities to achieve economically feasible manufacturing. Therefore, the fossil energy usage involved in utilities and capital costs are often high for chemical production. Furthermore, safety, sustainability, and waste disposal defy current trends (Burk & Van Dien, 2016).

However, there has been a recent emergence and expansion of the bioeconomy, which refers to economic activity involving the use of biotechnological principles to produce goods and provide services. To support the dream of a sustainable future, govern-

ment policies have been enticed toward bio-based industries. Biotechnology offers an alternative approach for the production of chemicals by using the biological pathways present in living organisms. Due to natural evolution over millions of years, microorganisms can create a range of materials that can be used for the production of chemicals, including renewable feedstocks (e.g., carbohydrates, glycerol, fatty acids, oils, etc.). One particular class of microorganism can also feed on single-carbon (C1) compounds (e.g., methane, carbon dioxide, carbon monoxide, and methanol) and mold them into useful chemicals. Such feedstock flexibility has a competitive edge against volatile and ecologically degrading petroleum-based raw materials.

A single cell of a living organism is like a miniature version of a chemical factory where all the chemical steps required in a chemical production plant occur within a micrometer-scale space. The use of such cells could reduce chemical plant space and become a more favorable alternative to large-scale production. Accordingly, capital expenses and energy costs tend to be significantly lower for bioprocesses. Notably, this option favors both economics and sustainability. A review by Weiss et al. (2012) compared 44 Life Cycle Assessment (LCA) studies of bio-based materials and showed that bio-based materials save 55±34 gigajoules of primary energy and 3±1 $CO_2$ greenhouse gas equivalents when compared to conventional materials. With recent advancements in genetic engineering, bioprocesses can be optimized to have significantly lower greenhouse gas emissions and fewer byproducts, thereby resulting in inexpensive waste disposal and a reduced carbon footprint due to lower energy demands. Regarding the safety of such processes, operations are performed at moderate temperatures and pressures with a reluctance to use toxic and flammable chemicals to ensure safety.

Since every coin has two sides, the flipside of biotechnology is its relatively low carbon yield, complexity of downstream separation process for product recovery and waste removal, and extensive R&D costs since this industry remains in the primitive stage of development.

Due to the expression of complicated metabolism networks in cells, the potential of biotechnology is being tapped by the search for natural products. Notably, 1,3-

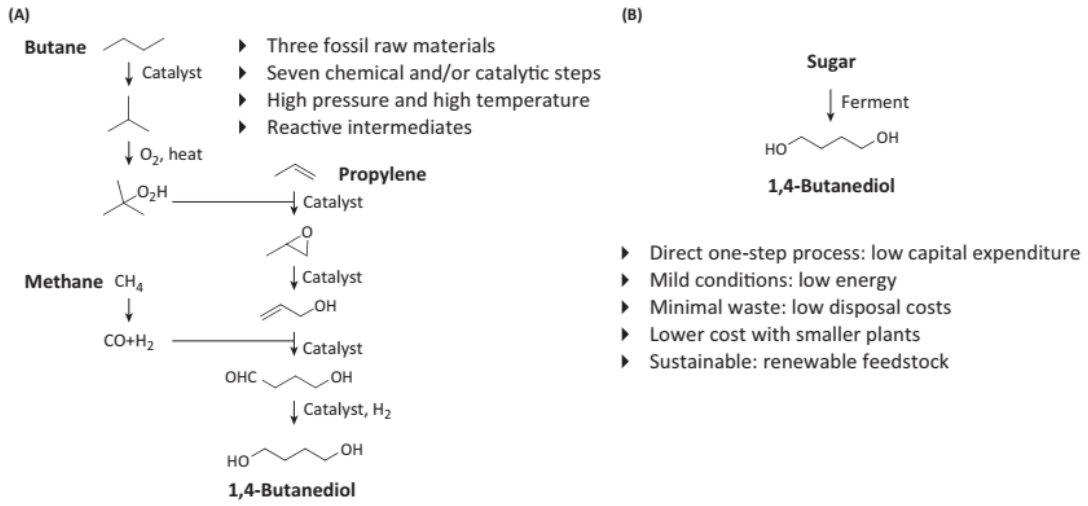



Figure 1.1: Comparison of 1,4-butanediol manufacturing methods. (A) Conventional petrochemical versus (B) bio-based processes (Burk & Van Dien, 2016).

propanediol (PDO) was the first chemical produced by fermentation using *Escherichia coli* as a host organism in a joint venture by DuPont and Tate&Lyle (Nakamura & Whited, 2003). An estimated 120,000 tons of 1,3-propanediol were produced in 2007. According to DuPont, the Bio-PDO process uses 40% less energy than conventional processes. This was soon followed by commercial production of 1,4-butanediol (BDO). Figure 1.1 compares commercially viable bioprocess with conventional chemical process for BDO production (Burk & Van Dien, 2016). Since then, many ventures have been undertaken to utilize biotechnology. Based on recent reports, the global market size for bio-based chemicals was valued at $8.81 billion in 2017 and is expected to witness a compound annual growth rate of 12.6% from 2018 to 2025.

Through an ongoing debate, biotechnology has made a convincing case and provided industry giants with incentives to pursue further bioprocess developments for natural products with applications in pharmaceuticals, food, petrochemicals, and fragrances, among others. However, developing a suitable alternative bioprocess is a tedious job. After narrowing down a product, moving from conception to materialization involves gene discovery and synthesis, strain and pathway optimization for the availability of precursors, and the removal of bottlenecks for the eventual development of a bioprocess yielding high titer and process scalability. In this thesis, we will focus on the

development of a bioprocess for terpenoid production.

## 1.2 Terpenoids

Terpenoids, also known as isoprenoids, are one of the largest families of natural products. Over 80,000 different terpenoids have been identified in nearly all life forms. Although the majority of terpenoids have been found in plants, they also occur in insects (Laurent et al., 2003), bacteria (Yamada et al., 2015), and fungi (Quin et al., 2014). In plants, they are produced as primary metabolites (e.g., sterols, carotenoids, quinones, and hormones) and secondary metabolites designated with the responsibility of plant defense and communication (Lange et al., 2000). Animal-derived terpenoids are involved in cell membrane formation, glycoprotein biosynthesis, and intracellular electron transport in the form of cholesterol, dolichol, and ubiquinone, among others (Goldstein & Brown, 1990). This makes them compounds of interest with useful applications in pharmaceuticals (e.g., artemisinin), flavoring (e.g., limonene), fragrances (e.g., patchoulol), biofuels (e.g., bisabolene), fuel additives, and agricultural pesticides (e.g. linalool).

Terpenoids can be classified according to the number of C5 units involved during their synthesis: hemiterpenoids (C5), monoterpenoids (C10), sesquiterpenoids (C15), diterpenoids (C20), triterpenoids (C30), tetraterpenoids (C40), and polyterpenoids ($>$ C40) (Connolly & Hill, 1984). They are derived from corresponding terpenes via cytochrome P450s.

The skeletons of terpenoids are products of the condensation of multiple units of isopentenyl diphosphate (IPP) and its isomer dimethylallyl diphosphate (DMAPP), which are naturally generated by either the mevalonate (MVA) pathway or methylerythritol phosphate (MEP) pathway (also known as the non-mevalonate pathway). Both pathways are expressed in plants but differ in their localization. The MVA pathway enzymes are located in the cytosol, whereas the MEP pathway enzymes are found in the plastid. When it comes to microorganisms, the MVA pathway is prevalent in eukaryotes and archaea, while the MEP pathway is found in eubacteria (Kirby &

Keasling, 2009).

Figure 1.2 describes the biosynthesis of terpenes for both pathways (Daletos et al., 2020). IPP is sequentially condensed to DMAPP by isoprenyl diphosphate synthases to generate the linear precursors geranyl diphosphate (GPP; C10), farnesyl diphosphate (FPP; C15), geranylgeranyl diphosphate (GGPP; C20), and geranylfarnesyl diphosphate (GFPP; C25). They are modified by terpene synthases to produce monoterpenes, sesquiterpenes, diterpenes, and sesterterpenes. In a similar manner, triterpenes (C30) and tetraterpenes (C40) are derived from the condensation of two FPP or GGPP molecules, respectively, followed by the action of terpene synthases. Beyond the activity of terpene synthase, cytochrome P450s are involved in the development of biologically active compounds (i.e., terpenoids) with applications spanning multiple industries.

However, since several of these compounds are scarce in nature, their supply remains inconsistent. Most of these compounds are still extracted from natural resources (mostly plants), which becomes economically challenging due to the low yield and long life cycle of plants. For example, the anticancer drug Taxol® is derived from diterpene paclitaxel extracted from the bark of Pacific yew trees. In the 1990s, it was estimated that 2–3 million Pacific yew (*Taxus brevifolia*) trees were required to meet the demand for cancer treatment in the USA alone (Suffness, 1995). Additionally, plant growth depends on environmental factors such as geographical location, which can cause variance in product quality and quantity. Chemical synthesis could be sought as an alternative for many of these terpenoids. However, as mentioned in the previous section, this process can be very energy demanding. Additionally, the complexity and inefficiency of production due to the stereoselectivity of the product application results in the formation of byproducts that cannot be used for the intended purpose. In this grim scenario, production in a microbial host stands out as a plausible alternative. Microbial hosts can start*de novo* terpenoid synthesis from simple carbon sources such as glucose, glycerol, and galactose. Like all biotechnological processes, it comes with the advantages of limited spatial requirements, stereoselectivity due to the activity of

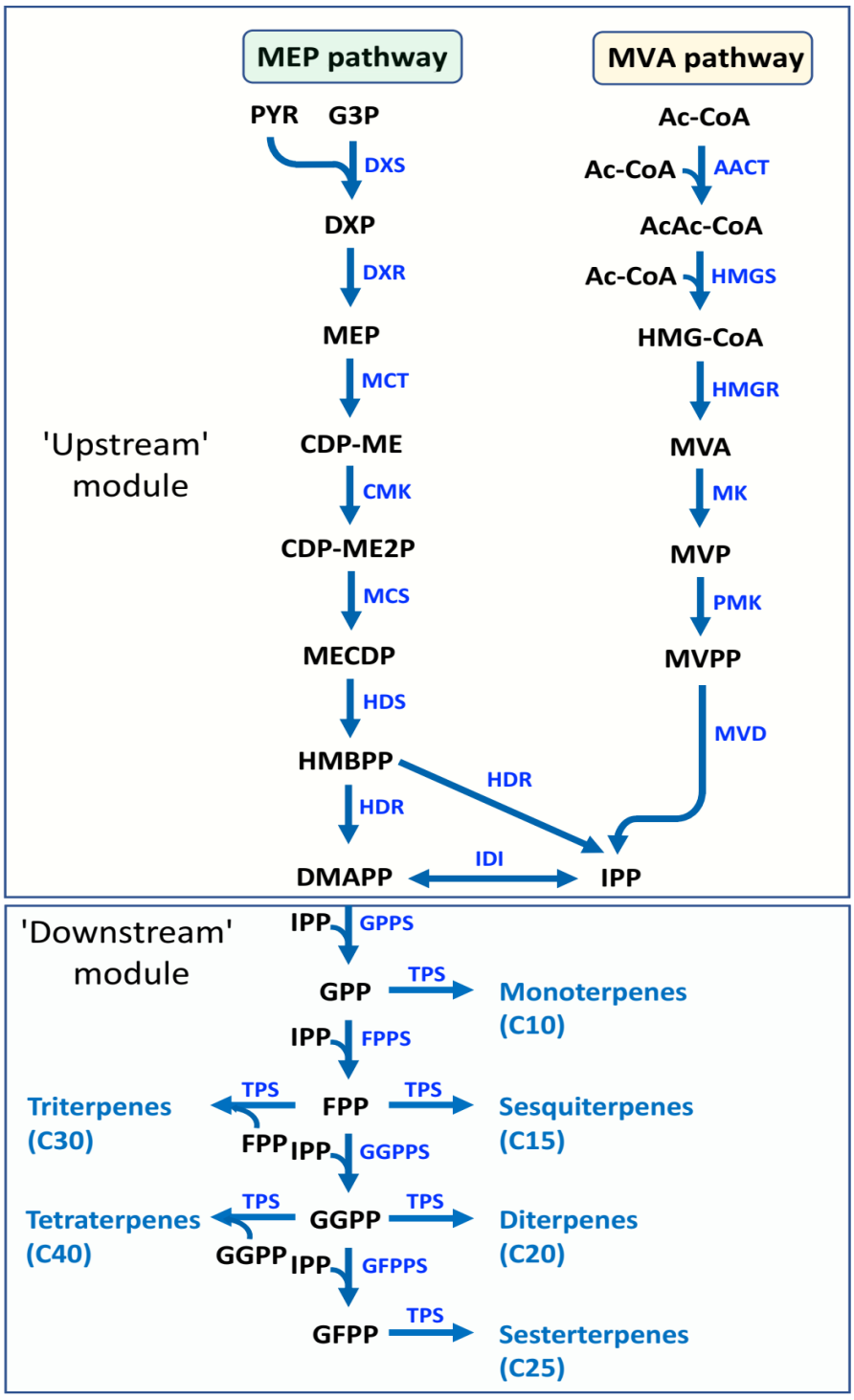



Figure 1.2: Production of terpenes via the MVA and MEP pathways. DMAPP and IPP condense together to form precursors for terpenes. Terpenes are further modified via cytochrome P450 to form corresponding terpenoids (Daletos et al., 2020).

natural enzymes, and sustainability as a result of the lower carbon footprint. Some challenges associated with the development of a microbial host for plant isoprenoid biosynthesis are presented as follows.

- Engineering of isoprenoid precursor pathways and gene discovery
- Incomplete characterization of the genetic circuit adding uncertainty
- The modification of terpenes and terpenoids to more bioactive molecules remains inefficient
- Accumulation of intermediates
- Inadequate supply of precursors (i.e., acetyl-CoA, pyruvate, and G3-P)

Amid the pros and cons, one important question that requires consideration is the selection of a pathway for terpenoid production.

## 1.3 MEP Pathway

The MVA pathway was discovered in the 1950s as a prominent pathway for terpenoid production (Little & Bloch, 1950). As seen in Figure 1.2, it begins with two molecules of acetyl-CoA condensed to acetoacetyl-CoA and another molecule of CoA via acetyl-CoA acetyltransferase (AACT). Acetoacetyl-CoA and another molecule of acetyl-CoA are then condensed by hydroxymethylglutaryl-CoA synthase (HMGS) to form 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA), which is further reduced by 3-hydroxy-3-methylglutaryl-CoA reductase (HMGR) to produce mevalonate. This is followed by phosphorylation via mevalonate kinase (MK) and phosphomevalonate kinase (PMK) on mevalonate and mevalonate-5-phosphate, sequentially. Mevalonate-5-phosphate is decarboxylated by diphosphomevalonate decarboxylase (MVD) to produce IPP, which is then isomerized to DMAPP by isopentenyl diphosphate isomerase (IDI) (Kaneda et al., 2001). In seeking to improve existing knowledge of terpenoid synthesis, researchers

conducted further investigations of IPP synthesis and discovered multiple methods of terpenoid formation in plants. In one of these studies, 14C-labeled mevalonate was rarely incorporated into the isoprenoids produced in maize seedling plastids, while the cytoplasmic sterols were quickly incorporated, thereby indicating an alternative sink of carbon synthesizing IPP (Goodwin, 1958). Further evidence of this was presented by a study in which 14C-labeled pyruvate and $CO_2$ were rapidly incorporated into a plastid-produced isoprenoid in *Sinapis alba* (Lütke-Brinkhaus & Kleinig, 1987). However, the existing MVA pathway had no expressed route to involve the assimilation of pyruvate to produce IPP. A follow-up on these findings led to the uncovering of the MEP pathway by Rohmer et al. (1993) in 1993. Earlier work has already shown that the production of pyridoxol (vitamin B6) stems from 1-deoxy-D-xylulose (Hill et al., 1989). A groundbreaking discovery of the incorporation of isotope-labeled 1-deoxy-D-xylulose into a ubiquinone of *E. coli* led to accelerated efforts in MEP pathway research.

The use of $^{13}$C-labeled glucose as a starting substrate for isoprenoid synthesis has significantly contributed to understanding the differential contributions of the MVA and MEP pathways in plants (Eisenreich et al., 2004). These studies have also been used to understand the reaction mechanisms of various steps in the pathways. Monoterpenes and diterpenes are predominantly biosynthesized via the MEP pathway (Eisenreich et al., 1998). This expands the scope of using synthetic biology for the production of various terpenoids using the MEP pathway. Additionally, carotenoids, which play a crucial role in photosynthetic organisms, are derived from the methyl erythritol phosphate pathway (Harker & Bramley, 1999). Other plant-based natural product derivatives of the MEP pathway include taxoids, alkaloids, essential oils, anthraquinones, and cannabinoids, among others (Ramawat & Mérillon, 2013, Sacchettini & Poulter, 1997). In the MEP pathway, pyruvate and glyceraldehyde-3-phosphate (G3P/GA-3P) are condensed by 1-deoxy-D-xylulose-5-phosphate synthase (DXS) to form 1-deoxy-D-xylulose-5-phosphate (DXP). Pyruvate is activated via a thiamine diphosphate cofactor and condensed to DXP after decarboxylation (Sprenger et al., 1997). DXS

is related to transketolases and E1 subunit of pyruvate dehydrogenase complex but belongs to novel group of carboligases (Schörken & Sprenger, 1998). Overexpression of DXS results in increased terpenoid production but also reduces the growth rate of the whole cell (Kim & Keasling, 2001). Further studies have shown that besides pyruvate, hydroxypyruvate and 2-oxybutyrate could also be a substrate for the DXS enzyme (Schürmann et al., 2002). In plants, DXS is a rate-limiting enzyme for the MEP pathway (Eubanks & Poulter, 2003). By using NADPH, DXP is reduced by 1-deoxy-D-xylulose-5-phosphate reductoisomerase (DXR) to form MEP, which is further catalyzed by 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase (MCT) to generate 4-cytidine 5-diphospho-2-C-methyl-D-erythritol (CD-ME). DXR was first isolated from *E. coli*, isomerizes DXP to 2-C-methyl-D erythritol 4-phosphate, and works with the cofactors $Mn^{2+}$ and $Mg^{2+}$ (S. Takahashi et al., 1998). MCT transferred cytidine triphosphate to MEP to generate CD-ME and pyrophosphate (Rohdich et al., 1999). CD-ME is phosphorylated by 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (CMK) to form 4-diphosphocytidyl-2-C-methyl-D-erythritol 2-phosphate (CDP-ME2P) using ATP as a co-substrate and bivalent magnesium ion as cofactor (Lüttgen et al., 2000). CMK is the only enzyme in the MEP pathway that is directly dependent on ATP as a substrate. CDP-ME2P undergoes cyclization and ring opening by the action of 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (MCS) and 4-hydroxy-3-methylbut-2-enyl- diphosphate synthase (HDS/IspG) to form 2-C-methyl-D-erythritol 2,4-cyclodiphosphate (MECDP/ MEcPP) and (E)-4-Hydroxy-3-methyl-but-2-enyl pyrophosphate (HMBPP), respectively. Notably, the MCS enzyme releases cytidine monophosphate during cyclization (Herz et al., 2000). The ring-opening reaction catalyzed by IspG uses reduction equivalents supplied by NADPH using flavodoxin reductase and the flavodoxin shuttle system (Seemann et al., 2002). HMB-PP reductase (HDR/IspH) acts on HMBPP to form DMAPP and IPP, unlike isomerization in the MVA pathway. IPP and DMAPP are formed in a molar ratio of 5:1 by IspH (Rohdich et al., 2002). IspH also uses electrons from NADPH transferred via the flavodoxin shuttle to catalyze the reaction. Although IDI can isomerize IPP

to DMAPP, its activity is not crucial when formed using the MEP pathway (Hahn et al., 1999). DXR, IspG, and IspH are the most challenging enzymes for the MEP pathway. DXR can result in two different products: MEP and 1-deoxy-L-ribulose-5-phosphate (an undesired product) (Wong & Cox, 2007). Since IspG and IspH are iron-sulfur cluster-containing proteins, they are sensitive to oxidative stress ((Giner & Jaun, 1998), Eisenreich et al., 2004). Hence, these three enzymes pose a challenge to terpenoid synthesis via the MEP pathway.

The stoichiometric balance for both pathways starting from glucose shows that carbon yield is 0.79 mol IPP/mol of glucose, generating 2.5 mol of $CO_2$ per mol of IPP via the MEP pathway. There are no redox molecules (e.g., NADH) produced in the MEP pathway. While the MVA pathway has a yield of 0.66 mol IPP/mole of glucose with the generation of 4 mol of $CO_2$ and 4 mol of NADH (Yadav et al., 2012). Despite the MEP pathway being carbon efficient, the MVA pathway is energy efficient. This results in the formation of excess NADH, which can reduce the amount of ATP required for IPP production via the MVA pathway. The MEP and MVA pathways have been targets of metabolic engineering efforts that aimed to enhance IPP production in host microorganisms. Production via the MVA pathway is more successful when compared to that of the MEP pathway due to the complexity of regulations involved in the MEP pathway. This has resulted in a head start for R&D efforts aimed at unraveling the kinetics of the MVA pathway. However, each pathway has potential depending on the product and microbial host involved.

## 1.4 Microbial hosts

To date, *E. coli* and *Saccharomyces cerevisiae* have been extensively studied as host microorganisms for terpenoid production. This preference can mainly be attributed to extensive knowledge of the genomics, genetic engineering, metabolism, and cell biology of these two microbes, which was already available two decades ago when metabolic engineering for terpenoid production was still in its infancy. Additionally,

rapid growth and relatively simple cultivation conditions help the operability of these processes, which is critical when selecting a host for production on an industrial scale. Therefore, especially for bulk chemical production, in which the efficient utilization of each supplied carbon atom is of high economic importance, these two organisms remain the first choice for production. Successful commercialization has been accomplished for artemisinic acid and $\beta$-farnesene with titers >25 g/L (Paddon et al., 2013) and >130 g/L (Meadows et al., 2016), respectively in *S. cerevisiae*. This was achieved by advanced genetic engineering efforts that involved debottlenecking the MVA pathway, addressing the toxicity of the intermediates, and increasing the acetyl-CoA supply. Due to fast-paced strain engineering working in concert with high-throughput screening, other alternative hosts have been showing signs of potential for isoprenoid synthesis. Despite this, a trade-off must be made between toxicity and high yield for certain terpenoids (Moser & Pichler, 2019). For example, monoterpenoids seem to have a toxic impact on certain physiological attributes (e.g., membrane integrity) of production hosts (Sikkema et al., 1995). This could be circumvented by using biphasic fermentation, which involves an added cost related to downstream separation. As an alternative, *Pseudomonas putida* offers high tolerance to monoterpenoids and engineered strains have been successfully applied for the *de novo* production of geranic acid to yield 193 mg/L (Mi et al., 2014). The production of heavier terpenoids in non-carotenogenic hosts requires the heterologous expression of a phytoene synthase, which can be a problem with *E. coli* and *S. cerevisiae*. On the other hand, *Yarrowia lipolytica* has shown higher yields for different carotenoids and can form large lipid bodies in which high amounts of hydrophobic compounds (including carotenoids) can be stored (Matthäus et al., 2014). Thus, depending on the product, a thorough investigation must be conducted to choose an appropriate host for terpenoid production. Some of the reported values for different hosts are presented in Figure 1.3.


Figure 1.3: Highest reported value for each terpenoid class produced by different microbial hosts. Mode of cultivation: shake flask (sf), bioreactor (br), microtiter plate (mtp). Values were taken as stated in the literature, preferably as a titer (mg/L or g/L); otherwise, as a space time yield (mg/L/d) or specific yield (mg/g) (Moser & Pichler, 2019).

Irrespective of the host, both of these pathways have certain limitations that must be overcome for them to perform optimally. Regarding the MEP pathway, imbalances in the supply of G3-P and pyruvate can create bottlenecks that lead to decreased pathway performance. Moreover, some issues arise with the cofactors involved in the pathway reaction. Flavodoxin reductase acts in coordination with iron-sulfur clusters containing IspG and IspH to form IPP and DMAPP. These clusters are sensitive to oxygen levels and can be inactivated these levels exceed a certain threshold, which leads to the accumulation of intermediates such as MEcPP (Moser & Pichler, 2019). Pathway intermediates or downstream products have been shown to inhibit front-end enzymes (e.g., IPP pools inhibit 1-deoxy-D-xylulose 5-phosphate synthase). Similar observations were made for the MVA pathway, where IPP, acetoacetyl-CoA, and HMG-CoA inhibited HMG-CoA synthase. With both pathways requiring redox cofactors and energy bonds to complete the reaction, complexity is introduced in the system since there are multiple competing reactions occurring simultaneously in the cell. These complex regulations can hinder attempts to upregulate either pathway (Chatzivasileiou et al., 2019). After altering the upstream module, the expression of terpene synthase and cytochrome P450s can be a challenge for most microorganisms. Several metabolic and protein engineering efforts have been made to optimize the production of terpenoids. This thesis revolves around unlocking the potential of *E. coli* as a host for the MEP pathway using optimized process conditions. In doing so, we have considered all limitations of the pathway and attempted to resolve them using process parameter optimization rather than metabolic engineering of the strain.

## 1.5 *Escherichia coli*

In 1884, German microbiologist and pediatrician Theodor Escherich studied infant gut microbes and their role in digestion and disease (Escherich, 1988). During this study, he discovered a fast-growing bacterium community that he called *Bacterium coli*, which later came to be known as *Escherichia coli*. The robustness and adaptability of *E. coli* has made it a center of microbiology efforts starting from the early 20th century.

As a result, it became the first organism to have its genetic code, transcription, and translation identified to reveal information about its entire genome. As such, it has become a workhorse of industrial biotechnology. *E. coli* has been the organism of choice for industry and academic labs due to several facets that include the following (Blount, 2015).

- Rapid growth in chemically defined media
- High yield, resulting in the use of relatively inexpensive culture media
- Does not coagulate to form aggregates, resulting in easier handling
- Industrial scalability
- Extensive knowledge of its genetics and genomics
- Extensive knowledge of its transcriptome, proteome, and metabolome
- Biosafety of several strains

For all of its importance, *E. coli* is rather nondescriptive. It is a fairly typical gram-negative rod-shaped coliform bacterium belonging to the genus *Escherichia*. It typically measures approximately 2.0 $\mu$m in length and 0.25–1.0 $\mu$m in diameter, with a cell volume of 0.6–0.7 $\mu m^3$ (Rogers & Kadner, n.d.). Some of its strains possess flagella that are used to move through its environment, while hair-like pili allow it to attach to surfaces or other cells. Physiologically, it is a facultative anaerobic organism since it can grow effectively with or without oxygen. However, it cannot grow at extremes of temperature or pH. Phylogenetically, it belongs to a class of the Enterobacteriaceae and is closely related to pathogens such as *Salmonella*, *Serratia*, and *Yersinia pestis*. More importantly, sequencing has uncovered the remarkable flexibility and robustness of the *E. coli* genome, which contributes to its genetic and phenotypic diversity. In 2002, it was reported that three widely used strains of *E. coli* (i.e., K-12, O157:H7, and CFT073) share only 39.2% of their genes (Welch et al., 2002). The entire *E. coli* genome consists of 16000 genes, of which only 20% belong to the core genome (Blount,

2015). Notably, a standard lab strain has over 4.5 million base pairs and 4000 genes (Gordienko et al., 2013).

With the advent of synthetic biology, diverse phenotypes of *E. coli* have been identified and engineered. *E. coli* has been able to produce novel compounds and achieve commercial production through metabolic and bioprocess engineering efforts. Hence, it has been prioritized over other microbial hosts for the synthesis of natural products, which requires a thorough understanding of the genotype of a strain before it can be modified to express genes that can produce natural products (e.g., terpenoids). High cell densities have been achieved in cell suspensions with *E. coli* by identifying its nutrients requirements according to physiological demands. Coupled with its high growth rate, this strategy increases product yield and drives the process in an economically efficient manner. Given the advantages of carbon efficiency in the MEP pathway and its native expression in *E. coli*, we decided to work toward exploiting the potential of this pair.

## 1.6 The MEP pathway in *E. coli*

As previously mentioned, the MVA pathway was always believed to be the sole source of terpenoid production until the MEP pathway was identified in the late 1990s. Hence, synthetic biology efforts always carried an advantage with the MVA pathway. Besides work to improve the synthesis of the isoprenoid framework for successful production, precursors (i.e., terpenes) must be altered to form bioactive molecules (i.e., terpenoids). This involves tailoring enzymes responsible for oxygenation, acylation, and glycosylation, etc. The oxygenation of isoprenes is catalyzed by cytochrome P450s and NADPH-dependent cytochrome P450 reductases (CPRs) (Renault et al., 2014). Their expression is a major challenge in *E. coli*, which lacks a natural system of P450 enzymes. However, *E. coli* flavodoxin reductase enzyme can serve as an alternative reductase partner for CPR in some cases (Tian et al., 2019).

Cognizant of these challenges, several groups have been focusing on the optimization of the MEP pathway.Ajikumar et al. (2010) optimized the IPP and DMAPP supply using

a multivariate-modular approach to taxadiene synthesis involving the installation of a flux balance between different modules of the MEP skeleton. Through this approach, 1 g/L of taxadiene was produced with *E. coli* in a fed-batch culture, representing a 15000-fold increase over the control strain. The functional expression of levopimaradiene synthase (LPS) in *E. coli* was only shown to generate minute but significant amounts of levopimaradiene (0.15 mg/l), a terpene precursor of the ginkgolides. With the goal of improving levopimaradiene production, geranylgeranyl diphosphate synthase (GGPPS) and LPS were engineered to improve their catalytic function, which, in combination with 10-fold overexpression of MEP genes, yielded 700 mg/l of levopimaradiene in a bench-scale bioreactor (Leonard et al., 2010).

An equimolar supply of G3-P and pyruvate is required for DMAPP and IPP generation via the MEP pathway. Improving the supply of these gateway molecules has been a focal point for chassis improvement efforts. The enhanced production of lycopene was achieved by overexpression of phosphoenolpyruvate carboxykinase (*pck*) and phosphoenolpyruvate synthase (PPS) as well as the deletion of pyruvate kinase (PykF and PykA). This prevented the accumulation of pyruvate, thereby establishing the balance between pyruvate and G3P while successfully improving the production of IPP and its conversion to lycopene in *E. coli* (Farmer & Liao, 2001). This highlights the intricacies involved in MEP pathway optimization, which cannot be resolved by enhancing the precursor supply alone since it must also be equipoised.

In another study, five MEP feeding modules derived from four glycolytic pathways in *E. coli*, which are the Embden–Meyerhof pathway (EMP), Entner–Doudoroff pathway (EDP), pentose phosphate pathway, and Dahms pathway, respectively, were tested for terpenoid production. It was observed that the module containing the EDP was more productive than that containing the EMP (Liu et al., 2013). This has provided a new scope for reworking the chassis of glycolysis in *E. coli*. Regarding MEP pathway enzymes, DXS, IDI, IspD, and IspF have been identified as culprits for rate limitation. Kinetic parameters for DXS $k_{cat}/K_M$ value are substantially lower than for other enzymes of the pathway (Kuzuyama et al., 2000). Thus, the heterologous expression of

enzymes derived from other hosts in *E. coli* has been examined alongside protein engineering efforts. For example, the heterologous expression of *dxs* from *Bacillus subtilis* improved IPP, DMAPP, and -carotene production in *E. coli* (Leonard et al., 2010). NADPH supply is critical for the second enzyme in the MEP pathway, DXR, which could be improved by rewiring the glycolysis pathway and improving the process parameterization for cultivation. In this thesis, we have focused on the optimization of NADPH supply to explore the potential of our strains. Simultaneously, to conserve the activity of IspG and IspH enzyme, it is very important to maintain environmental oxygen below the critical threshold to avoid the degradation of iron-sulfur clusters acting as cofactors. Thus, the installation of an optimal dissolved oxygen level is mandatory when conducting terpenoid production via the MEP pathway. During this thesis, we worked on the production of amorpha-4,11-diene (a precursor for artemisinin) and a blend of valencene and nootkatone (i.e., EGN).

## 1.7 Artemisinin

Malaria is a life-threatening disease caused by parasites transmitted from mosquitoes of the genus *Anopheles*. In 2018, there were 228 million cases of malaria resulting in an estimated 40,500 deaths worldwide (WHO, 2019). Malaria is caused by the bite of female *Anopheles* mosquitoes infected by *Plasmodium* parasites. There are five parasite species that cause malaria in humans, and two of these species—*P. falciparum* and *P. vivax* —cause the most dangerous form of malaria. The early diagnosis and treatment of malaria reduces disease severity and prevents death. Several drugs have been developed for the treatment of malaria. However, resistance among the parasites has made these treatments ineffective. For example, *P. falciparum* has become resistant to chloroquine. A similar observation has been made for artemisinin. However, a combination of antimalarial drugs is proven to be effective in the treatment of malaria. The best available treatment, particularly for *P. falciparum* malaria, is artemisinin-based combination therapy (ACT) and has been recommended by the World Health Organization (WHO.) Fast-acting artemisinin-based

compounds are combined with a drug from a different class. Companion drugs include lumefantrine, mefloquine, amodiaquine, sulfadoxine/pyrimethamine, piperaquine, and chlorproguanil/dapsone. Artemisinin derivatives serving as a complementary part for ACTs include dihydroartemisinin, artesunate, and artemether. By the end of 2016, the use of ACTs had been adopted as the first-line treatment policy in 80 countries. Thus, the demand for artemisinin has continued to grow over the past decade. An increased procurement rate of up to 31% has been recorded. Most of the distribution has been occurring in African countries where the incidence of malaria has been higher than anywhere else in the world (World Health Organization, 2018). As a result, the cost of this treatment has been a major concern and efforts are being made to address this.

Artemisinin is derived from the extracts of sweet wormwood (*Artemisia annua*). Artemisinin is a sesquiterpenoid and its efficacy also extends to parasitic infections such as schistosomiasis. It was first highlighted as an antimalarial drug in Chinese traditional literature (i.e., Ge Hong's Zhouhou Beiji Fang Handbook of Prescriptions for Emergency). It has subsequently been included in Chinese medical writings (e.g., Bencao Gangmu) (Krishna et al., 2008). The discovery of artemisinin as an antimalarial drug was an endeavor of 'Project 523', which was a collaboration by Chinese scientists in the 20th century (Zhang, 2005) that resulted in the development of dihydroartemisinin, artemether, and artesunate. These derivatives were identified as crucial antimalarial drugs. The current supply of artemisinin is primarily dispensed upon extraction from cultivated sweet wormwood plants. The synthesis of artemisinin, an isoprenoid molecule that contains 15 carbon atoms and a lactone endoperoxide, is depicted in Figure 1.5. *A. annua* generates sugar via photosynthesis, directing it toward the glycolysis pathway to form acetyl-CoA. Acetyl-CoA makes headway to the MVA pathway to generate IPP, which is converted into an intermediate for sesquiterpene production, FPP. The first directed step of artemisinin biosynthesis is the conversion of FPP to amorphadiene, a sesquiterpene, by amorphadiene synthase (ADS). Amorphadiene is oxidized to artemisinic acid and dihydroartemisinic acid. Dihydroartemisinic acid is a

direct precursor to artemisinin that is converted via sunlight to the desired artemisinin (Paddon & Keasling, 2014).

Notably, the conventional method of artemisinin production has been at the mercy of fickle agriculture market dynamics. Figure 1.4 explains the market dynamics of artemisinin until 2015 (Tsuruta et al., 2009). Uncertainty in supply is caused by an over-dependence on environmental factors that require predictions and planning. Yields of artemisinin from *A. annua* are very low and a substantial increase in the cultivation of plants is required (Hale et al., 2007). Attempts are being made to move to new methods of increasing the artemisinin supply in order to stabilize the supply chain and ultimately increase access to ACTs in developing countries. Given problems associated with the complexity and lower yields of chemical synthesis, alternative artemisinin production methods using microbes have been sought to maintain a consistent supply.

As an initiative toward the market stabilization of artemisinin supply, a project was funded by a grant from the Bill and Melinda Gates Foundation in 2004. This was led by a partnership between the University of California (Berkeley, USA), Amyris Inc., and the Institute for OneWorld Health. A semi-synthetic process was developed by Amyris Inc. for the production of artemisinic acid in yeast, which could be further converted into artemisinin via chemical synthesis.

Amorpha-4,11-diene is an important precursor for the semi-synthetic production of artemisinin. In microbes, genetic modification has been implemented to produce artemisinic acid from amorpha-4,11-diene. Beyond this, conversion is achieved through a chemical pathway by converting artemisinic acid to dihydroartemisnic acid followed by esterification, which has very low yield due to the formation of an isomer byproduct that can be further converted into artemisinin. *E. coli* was used as a host for the synthesis of amorpha-4,11-diene via heterologous expression of the MVA pathway. Using this method, titers of up to 27 g/L were achieved (Tsuruta et al., 2009). However, the oxidation of amorpha-4,11-diene to artemisinic acid requires the heterologous expression of cytochrome P450, which has been a chronic challenge for *E. coli*. Thus,

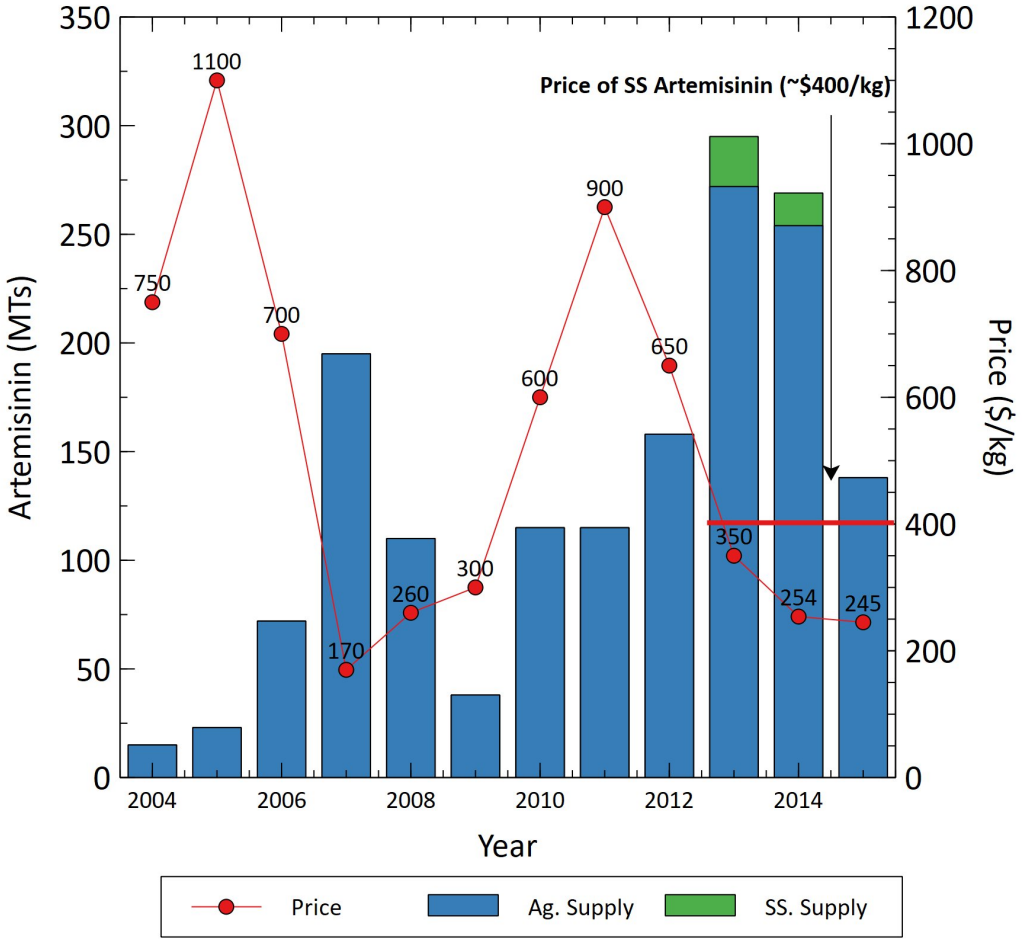



Figure 1.4: Market dynamics of artemisinin from the year 2004 to 2015. Blue bars represent agricultural supply while the green bars represent production using a semi-synthetic process. The market price of artemisinin is indicated by the red line with dots. The cost of semisynthetic artemisinin production is approximately \$400/kg, which is higher than the market price.

*S. cerevisiae*, which already has P450 present along with a cognate reductase module, has been used as a preferred host. An *S. cerevisiae* strain capable of producing 40 g/L of amorpha-4,11-diene was developed by Amyris Inc.. However, the conversion of amorphadiene to artemisinic acid was poor when expressed in yeast (Paddon et al., 2013). When subjected to further improvements of the P450/CPR assembly, the expression of reductase AaCPR was decreased to balance the CYP71AV1 and AaCPR interaction, and an artemisinic acid titer of 25 g/L was obtained.

However, the cost of semi-synthetic artemisinin production is approximately $400/kg against a $181/kg selling price. Hence, such production was halted in 2015 according to an article published in 2016 (Peplow, 2016). Hence, new alternatives are being sought for process development. We plan to take advantage of the high carbon yield of the MEP pathway to help reduce this cost via the production of its precursors in *E. coli* through a native MEP pathway. The current process involves developing a platform for the production of amorpha-4,11-diene and then further extending it to produce artemisinic acid.

## 1.8 EGN

As part of this thesis, we worked on the production of an EGN blend in *E. coli* via the MEP pathway. This was a mixture of valencene, nootkatol, and nootkatone.

Nootkatone is a sesquiterpenoid with a characteristic grapefruit-like flavor and is the main chemical component of smell and flavor in grapefruit (Haring et al., 1972). Conventionally, nootkatone has been extracted from grapefruit (Del Rio et al., 1992). Like artemisinin, the agro-based production of nootkatone fluctuates and is limited by geographical and environmental factors. Hence, alternatives are being sought for its production. Synthetic nootkatone can be produced from valencene via oxidation. The chemical process of producing nootkatone from valencene involves the use of tert-butyl chromate (Hunter & Brogden Jr, 1965). This process has a yield of approximately 76%. However, ter-butyl chromate is carcinogenic and the process has a running risk of explosion if not conducted at room temperature. Wilson III & Shaw (1978) developed a

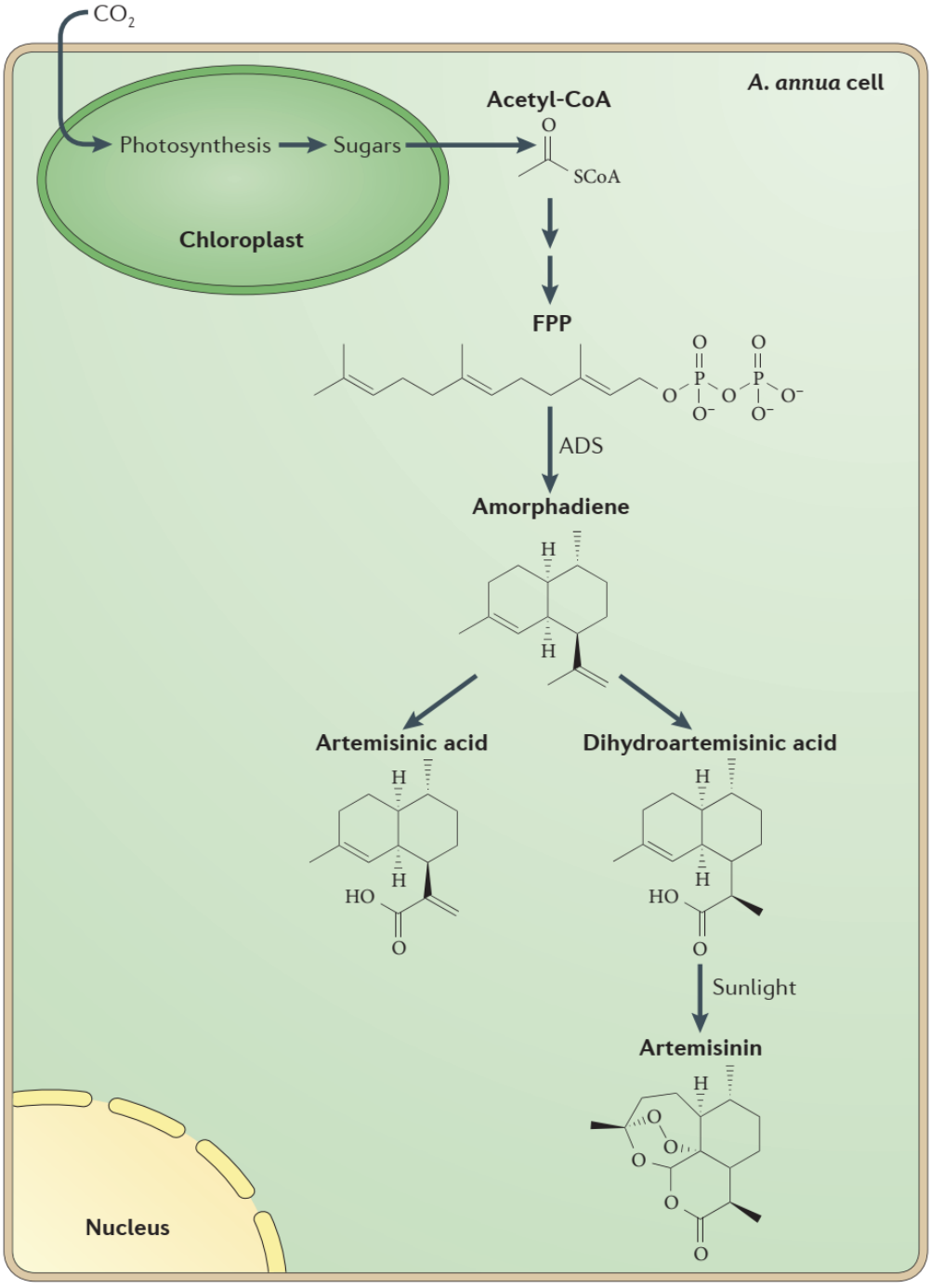



Figure 1.5: Artemisinin biosynthesis pathway in the plant *Artemisia annua*. Sugars that are produced by photosynthesis in plant chloroplasts are converted to acetyl-CoA in the cytosol. The two-carbon acetyl-CoA feeds into the cytosolic mevalonate pathway, which produces FPP. FPP is then converted to 15-carbon sesquiterpene amorphadiene by the enzyme amorpha-4,11-diene synthase (*ADS*). Amorpha-4,11-diene is then enzymatically oxidized to either artemisinic acid or dihydroartemisinic acid. Dihydroartemisinic acid is the precursor of artemisinin and is thought to be converted to artemisinin in a non-enzymatic photochemical reaction that is stimulated by sunlight (Paddon & Keasling, 2014).

three-step process using ter-butyl peracetate as an oxidizing agent to convert valencene to nootkatone in a less hazardous process. However, chemical methods come with the challenges of low yields, unwanted byproducts, and a higher carbon footprint. Hence, biotechnological alternatives are being investigated. The biosynthetic path for converting valencene to nootkatone involves regioselective allylic hydroxylation to form nootkatol followed by oxidation to nootkatone. Figure 1.6 depicts a two-step conversion process catalyzed by valencene synthase to convert FPP to valencene and convert multifunctional cytochrome P450 enzyme hydroxylation and oxidizing valence to nootkatone (Cankar et al., 2014). Several cytochrome P450s have been reported to catalyze valencene oxidation. Engineered cytochrome P450s from *P. putida* and *Bacillus megaterium* were shown to convert valencene to nootkatone (Seifert et al., 2009, Sowden et al., 2005). Valencene dioxygenase from *Pleurotus sapidus* also demonstrated the enzymatic conversion of valencene to nootkatone (Fraatz et al., 2009). Efforts are being made to express heterologous cytochrome P450s from plants in microorganisms. For example, P450 enzymes from *Cichorium intybus* have shown the ability to convert valencene to trans-nootkatol and nootkatone (de Kraker et al., 2003). Moreover, CYP71D51v2 from *Nicotiana tabacum* was reported to oxidize valencene to trans-nootkatol (Gavira et al., 2013). Additionally, Girhard et al. (2009) screened 125 P450 enzymes from a novel recombinant P450 library based on approximately 250 bacterial cytochrome P450 monooxygenases co-expressed with putidaredoxin reductase (PdR) and putidaredoxin (Pdx) in *E. coli*. They identified CYP109B1 from *B. subtilis* as the most effective enzyme for valencene oxidation. Also, Cankar et al. (2014) identified novel cytochrome P450 from *Cupressus nootkatensis* to catalyze valencene oxidation.

Valencene is a natural sesquiterpene found in various citrus species such as Valencia orange (Hunter & Brogden Jr, 1965). It has a sweet, citrus, and woody note and is used as an additive in the food industry. The extraction of valencene from citrus fruits has been deemed uneconomical. Hence, biosynthetic alternatives have attracted interest. Chen et al. (2019) reached 539.3 mg/L of valencene titer in a fed-batch process with *S. cerevisiae* by expressing valencene synthase. This process used the MVA path-

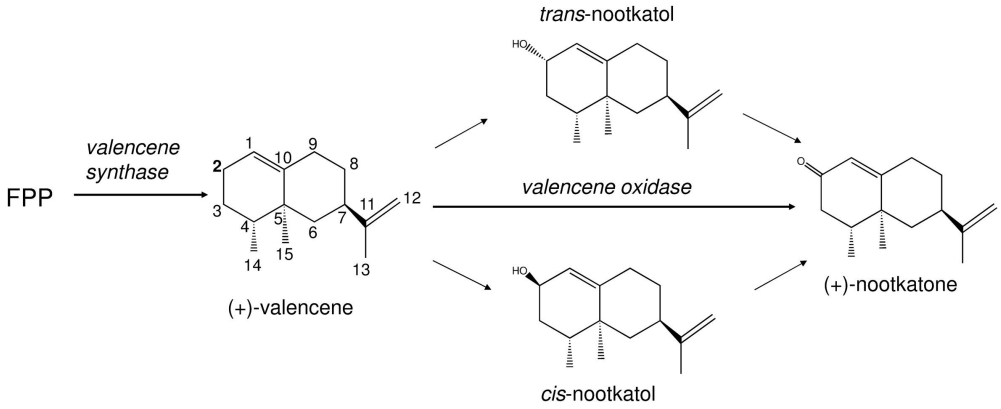



Figure 1.6: In the first enzymatic reaction, valencene formation from farnesyl pyrophosphate (FPP) is catalyzed by a valencene synthase. The two-step enzymatic conversion of valencene to nootkatone is proposed to proceed via hydroxylation at the 2-position of valencene, yielding trans- or cis-nootkatol intermediates followed by oxidation to nootkatone (Cankar et al., 2014).

way, which is native to yeast for the production of FPP and was further converted to valencene by valencene synthase from *C. nootkatensis.* Previous attempts at valencene production were made in *Rhodobacter sphaeroides* strains by heterologously expressing a mevalonate operon from *Paracoccus zeaxanthinifaciens* and valencene synthase (Beekwilder et al., 2014). Corynebacterium glutamicum was used a host to express valencene synthase from Nootka cypress to yield 2.41 mg/L of valencene.

In this thesis, part of the work consists of process development for the production of our EGN blend. Since the composition of the blend is deemed confidential for the project partner, we will not dive deep into the effect of the process conditions or the profile of the mixture. For the purpose of confidentiality, we will refer to the product as EGN and express the titer in arbitrary units (%AU/L).

## 1.9 Industrial production and scale-up

High cell density fermentation is targeted to improve the yield of product from a proof of concept from small to large scales. For commercialization, any process that can produce a designated product using an economically viable option is important. Bioprocess development can be elaborated from Figure 1.7. Micro-titer plates (MTPs)

and shake flasks remain the most popular option for high-throughput screening and preliminary process development studies (Wewetzer et al., 2015). Since only batch fermentation can be performed at this scale, maintaining a consistent process environment becomes an issue, and productivity is often compromised. Moreover, batch fermentation rarely faces substrate limitation. Thus, it can be difficult to replicate its performance on a larger scale. As such, the next stage of process development involves upgrading to the laboratory bioreactor scale. However, to address the potential issues of industrial production, it is important to run the process at the pilot plant scale since this resembles a production bioreactor. Process development at the lab scale involves the optimization of media and feeding strategies to selectively enhance production and maintain sustenance. Once a baseline process is developed, the platform could be used to evaluate strains. During this process, fermentation is studied biologically and mechanistically to pass information back to the strain development team. This process is repeated multiple times before a strain with optimal performance and robustness can be passed along to the pilot plant fermenters. Various cultivation parameters (e.g., media composition, pH, agitation, aeration, temperature, cell density, the concentration of inducers, induction time, and feeding strategies) affect the protein expression level based on the expression systems used (Gronemeyer et al., 2014). Thus, it is essential to evaluate each of the cultivation conditions for the expression of every recombinant protein and the development of effective bioprocesses that are scalable for commercialization. As a result, it is also important to identify the economics of the process.

Media represents close to 70% of the production cost. Hence, the development of an effective medium composition that includes all of the essential nutrients necessary for higher cell growth and protein productivity is a fundamental requirement. Various approaches are taken to define a cultivation medium. In doing so, it is important to identify the essential attributes of the process that can be related to performance. Cultivation media were previously developed using the traditional one factor at a time approach. With the advancement of statistical and data analysis, tools such as facto-

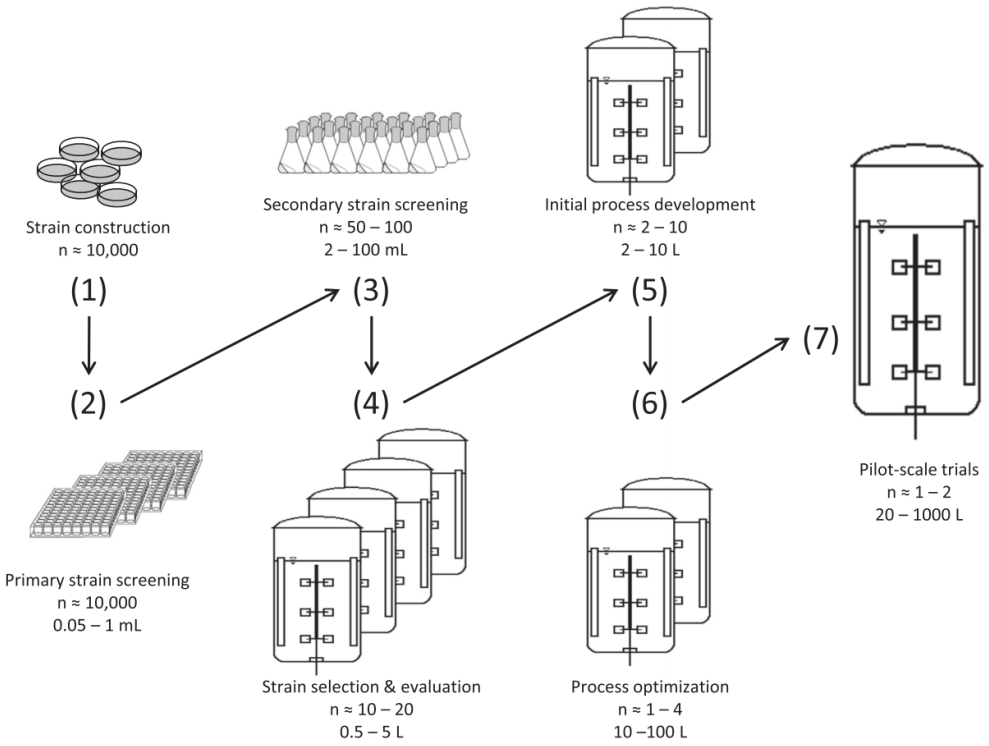



Figure 1.7: Strategy for bioprocess development. The conventional workflow from strain screening to production at the pilot plant scale (Wewetzer et al., 2015).

rial design, Plackett-Burman design, and dynamic simulations have been implemented. To prolong the process in order to achieve economically viable titers, it is necessary to maintain the supply of nutrients. This can be achieved by selecting the appropriate mode of cultivation. These modes can be identified as either batch, fed-batch, continuous, or perfusion. Batch mode involves the provision of all necessary nutrients in the initial batch medium. This is mainly restricted by limited biomass production. This mode is not very popular in the current biotechnology setting. However, for bioconversion processes, batch cultivation remains the most economical and feasible mode of cultivation. In fed-batch process, nutrients are fed during cultivation in addition to nutrients supplied in the initial base media. This system is useful when certain substrates are targeted to only act as a regulator switch for the production pathway. In perfusion culture, the medium is circulated through a growing culture to allow simultaneous waste removal and nutrient supply. While in a continuous or chemostat bioreactor culture, feed containing essential nutrients is fed in and product containing culture is continuously recovered. A chemostat is desired when certain variables

defining the dynamic state of the system are to be held constant. Notably, key performance indices often decide the cultivation mode. During process development, it is important to develop an interplay between metabolism and process parameters to reduce the number of experiments required for process optimization.

Once the process was ready for the lab-scale bioreactors, we scaled up the process to higher volumes. Scale-up comes with its own unique challenges. As a central problem in biotechnology, the scale-up of fermentation processes was first recognized and described during industrial penicillin production in the early 1940s (Schmidt, 2005). Fermentation scale-up is aimed at manufacturing larger product quantities targeting an increase in—or at least the consistency of—specific yields and product quality. To increase product yields and ensure consistent product quality, it is important to maintain optimum and homogenous reaction conditions to minimize microbial stress exposure and enhance metabolic robustness. However, changes in geometry can lead to less favorable mixing behavior and deterioration in physiological characteristics, which subsequently result in process variability, reduced biomass-specific yields, and more unwanted byproducts, thereby affecting the quantitative and qualitative attributes of the process. Due to inefficient mixing, pockets of high substrate and heat concentration are formed, which could affect the physiology of the cells.

One of the major issues related to scale-up is a reduction in mixing quality. This is characterized by long mixing times, which creates a zone of nutrient starvation in the bioreactor. This is exemplified by an inefficient Oxygen Transfer Rate (OTR), which is one of the most important variables in the aerobic processes. This can also act as a proxy for the mixing characteristics of the bioreactor. Air is typically sparged from the bottom of the tank in microbial fermentation, while the substrate is fed from the top. As a result, cells at the top of the fermenter are exposed to excess substrate concentrations and simultaneously suffer from oxygen limitation, whereas those at the bottom experience the opposite. This can result in the creation of a byproduct sink during cultivation. For example, excess glucose concentrations result in byproduct accumulation in the form of acetate, while simultaneous oxygen limita-

tion induces a mixed acid fermentation pathway for facultative anaerobes such as *E. coli* to form ethanol, lactate formate, and succinate, among other compounds (Bylund et al., 2000). Carbon dioxide solubility is affected by hydraulic pressure, which results in buffering capacity variation that ultimately affects pH correction. Variance in metabolic activities results in zonal overheating. When combined with carbon dioxide accumulation and toxic byproduct accumulation, this variance results in a high-stress environment for the cell. Notably, the subsequent activation of stress genes can only protect cells to a limited extent (Kvint et al., 2003). Cells can also undergo genotypic changes. Under environmental stress conditions, cells can be subjected to the unforced mutation of heterologous genes, which can significantly impact production. It has been reported that heterogeneity in mixing can affect protein formation within cells. One of the most important consequences of stress-induced metabolic shifts is the misincorporation of amino acids into native and recombinant proteins. Fenton et al. (1997) observed that changes in glycolysis, the citrate cycle, and pathways involved in anaerobic and mixed acid fermentation due to stress response led to shifts in the coupled amino acid biosynthetic pathways. This resulted in a shift in the amino acid pool, which led to inaccuracy in the inherent translation process and affected protein synthesis. Moreover, this led to an increase in byproducts at the expense of desired main product yields in large-scale fermentation.

## 1.10 Oxygen transfer rate

In aerobic processes, oxygen is an important nutrient used by microorganisms for growth, maintenance, and metabolite production. Since oxygen starvation can affect the performance index of an aerobic process (Çalik et al., 2004), it is important to ensure the adequate delivery of oxygen from a gas stream to the culture broth. This is determined by the diffusion characteristics of the system, which impact its mass transfer coefficient (a proxy for the mixing quality for the bioreactor). This makes diffusion characteristics among the most important scale-up parameters. Consequently,

the accurate estimation of OTR at different scales and under different operational conditions has a relevant role in the prediction of metabolic activity for both the growth and production of a culture. The oxygen mass transfer rate is proportional to the concentration gradient of oxygen between the gas/liquid phase interface and liquid phase, with volumetric mass transfer coefficient $k_La$ being the proportionality constant. Since concentration at the gas/liquid phase interface is governed by Henry's law, the maximum rate of oxygen transfer remains limited by the solubility of oxygen in the pertaining medium. Various parameters play a role in determining OTR, including nutrients, pressure, agitation, and aeration (see Figure 1.8). During cultivation, oxygen transport occurs simultaneously with the cellular respiration. Therefore, the biochemical states of cells interact with oxygen transfer to develop an equilibrium. The kinetics of the process are controlled by a rate-limiting step, which turns out to be OTR in most aerobic processes. Thus, it is important to address the mass transfer of oxygen in a bioreactor during scale-up.

A few different strategies could be employed in bioreactor scale-up, which are listed as follows (Garcia-Ochoa & Gomez, 2009).

- **Fundamental analysis:** This is based on the application of mathematical and physical models to describe the influence of operational conditions and the geometrical design of the bioreactor on the fluid dynamics of the bioreactor. While it obeys the fundamental laws of physics, it can be tiresome due to the tedious derivation of solutions.

- **Semi-fundamental analysis:** Semi-fundamental methods are those where simplified equations are applied to obtain a practical approximation of the fundamental laws. This analysis is not transferable and is only relevant to the system in question. However, it still involves complexities in development.

- **Dimensional analysis:** All the critical dimensionless groups of parameters are kept constant during the scale-up process. However, it is usually impossible to keep all dimensionless groups constant, which represents a drawback of this

approach.

- **Rule of thumb:** Based on the successful scaled-up process, a few of the criteria have been identified as crucial to match. This approach involves selecting a few of these criteria and scaling up the process parameters to match the values at different scales. Some of the commonly used criteria (e.g., power input per unit volume, $k_La$) are kept constant while scaling up.

The rule of thumb method is the most commonly used method in this context. The most commonly used scale-up criteria and their proportion of use in the fermentation industry are presented as follows: volumetric power input $P/V_r$, (30% of use); constant volumetric mass transfer coefficient $k_La$, (30%); constant impeller tip speed of the agitator or shear, (20%); and constant dissolved oxygen concentration $p_{O2}$, (20%) (Margaritis, 1978).

The different scale-up criteria normally result in entirely different process conditions on a production scale. Usually, it is impossible to maintain all of the parameters in the same ratio to one another. Hence, the use of criterion is pertinent to the individual bioprocess. Most of the aerobic bioprocesses are scaled up to match the mass transfer coefficient of the system (Herbst et al., 1992). The empirical correlation suggests that following the rule of matching the mass transfer of the system can cover some of the other rules of the thumb (e.g., power input, impeller tip speed, and aeration).

The empirical measurement of $k_La$ is performed using chemical and physical methods during cultivation. However, the measurement of $k_La$ is always a challenge in large bioreactors. Thus, empirical correlations are often used. The generalized form of the empirical equation is provided as follows.

$$k_La = \gamma(\frac{P}{V_r})^{\alpha}(v_g)^{\beta}(\mu_a)^{c} \tag{1.1}$$

Many models have been developed to determine the values of $\alpha$, $\beta$, and $\gamma$ in the equation. However, van't Riet's equation is most widely used for such calculations

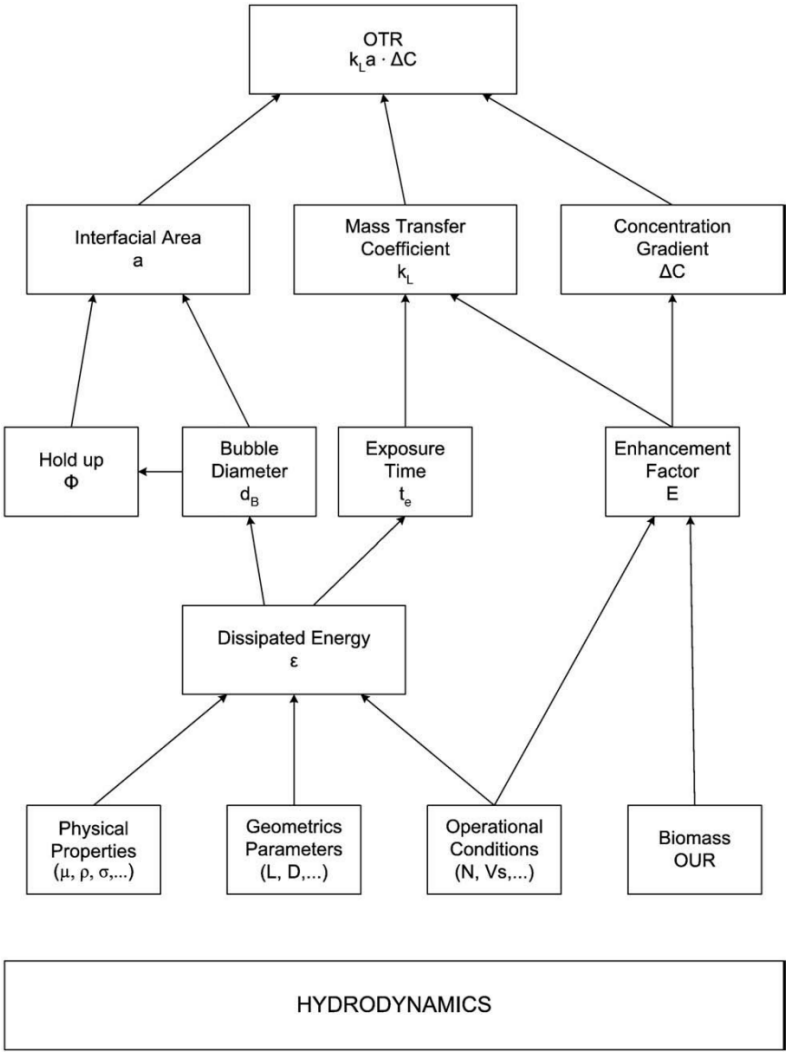



Figure 1.8: Relationship between OTR, volumetric mass transfer coefficient, and hydrodynamic parameters in a bioreactor (Garcia-Ochoa & Gomez, 2009).

(Van't Riet, 1979). For most of our scale-up calculations, we focused on the mass transfer coefficient, tip speed, and aeration while maintaining the same dissolved oxygen levels as small-scale fermenters.

## 1.11 Systems biology

Systems biology is a field of biochemical engineering that considers a holistic approach to deciphering the complexity of biological systems by understanding the network that constitutes a whole organism. It involves the collaboration of biology, computer science, engineering, bioinformatics, and physics to understand organisms' responses to varying conditions. Moreover, it has been widely used since the early 20th century.
To understand biology at the system level, we must examine the structure and dynamics of cellular and organismal function. Systems-level understanding can be derived using four key properties (Kitano, 2002):

- **System structures:** A network of gene interactions and metabolic pathways as well as their interactions that derive phenotypical and genotypical responses.
- **System dynamics:** The responses of systems under time varying conditions, including their distribution through networks.
- **Control method:** A physically-derived mechanism used to control the state of a cell.
- **Design method:** A principle-driven construction of the metabolic system.

Figure 1.9 depicts an efficient cycle of research hinged upon systems biology-based hypotheses instead of the trial and error method. This involves developing a computational model and then performing simulations to predict results, which is followed by conducting experiments. The results from experiments can be analyzed and the model can be further modified. This methodology implements several systems biology tools. Two of the major tools used in this thesis are omics and Flux Balance Analysis (FBA) using a stoichiometric model. Omics aims to perform the qualification and

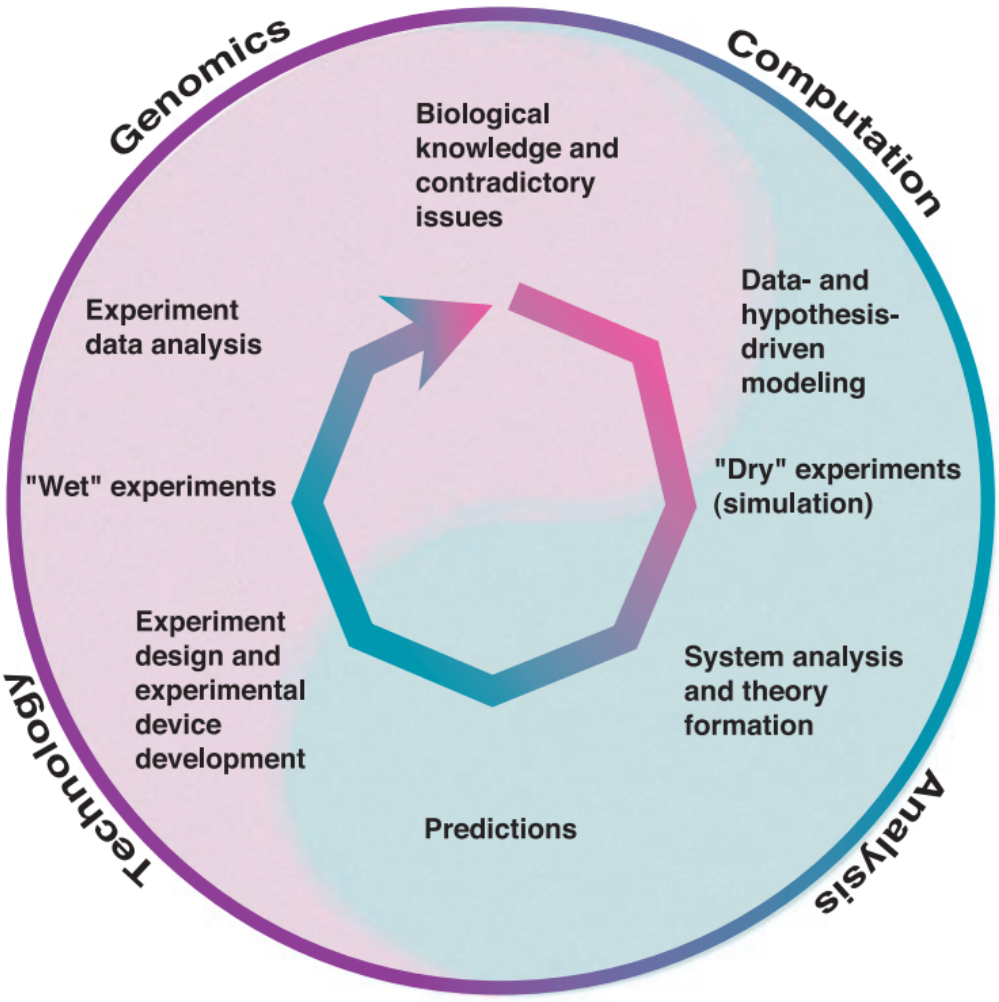



Figure 1.9: Cycle involving systems biology-based tools for research and development (Kitano, 2002).

quantification of pools of biological compounds to interpret the functions of the organism under study. Notably, we have implemented metabolomics and transcriptomics in this thesis. Steady-state analysis can be performed on a metabolic network via device constraints on certain reactions using FBA. Edwards et al. (2001) used FBA to predict the transition of the metabolic pathway in *E. coli* under different nutritional conditions. They used knowledge of the available metabolic network structure and confirmed the results experimentally. Likewise, we sought an in-depth understanding of our organism and performed bioprocess development based on that knowledge to reduce the number of experiments governed by trial and error methods.

# Chapter 2

# Objective and Scope

This study aimed to develop a platform to produce terpenoids in *E. coli* using the MEP pathway. Given the intricacies of bioprocess development related to the MEP pathway, we planned to identify the bottlenecks of the process and troubleshoot them by targeting them systematically. The study spans two products: EGN and amorpha-4,11-diene (AMD4,11). Due to a confidentiality clause with project partners Manus Bio Inc., we cannot disclose the actual name of the EGN product and the enzyme modification required downstream of the MEP pathway for EGN production. Since both products are developed using the same platform, we planned to achieve process transferability between the two products. The present study is divided into five sections, starting from the installation of baseline cultivation conditions for the EGN1 strain engineered to produce EGN. These conditions were then adapted to the AMD strain and we strove to establish the interplay between various process parameters and the dynamics of ATP and NADPH, which appear to drive the complexities of the process. Our aim was to develop an environment for the strain that would maintain an optimal balance between growth and production. In this thesis, we explore various methods of improving cofactor production to target enhanced MEP flux. On establishing the correct environment, the process was then transferred back to EGN production via the EGN2 strain. Since scaling up the process can influence the balance achieved, we also aimed to successfully scale up the process to a bioreactor with a 300L working volume. An outline of the thesis is as presented follows.

- Challenges in MEP process development manifested by EGN
- The fundamental role of proper glucose and oxygen supply for producing terpenoids in *E. coli* showcasing amorpha-4,11-diene
- The effect of Yeast Extract (YE) on MEP flux
- Amino acid metabolism and relevant effect on flux of the MEP pathway
- Transferability of the process

# Chapter 3

# Materials and Methods

In this chapter, the materials and methods used for this thesis are summarized. Part of this section has been published in Patil et al. (2020) and the corresponding references are provided.

## 3.1 Bacterial strains

Three strains of *E. coli* were cultivated as a part of this thesis. These strains, derived from *E. coli* MG1655, were engineered for enhanced flux through the MEP pathway toward the common terpenoid precursors IPP/DMAPP. Strain details are filed in Kumaran et al. (2019). One of the important chassis modifications of the parent strain relevant to this thesis is the knockout of the genes *pgi* (glucose-6-phosphate isomerase) and *ldhA* (D-lactate dehydrogenase), which are responsible for the inter-conversion of glucose-6-phosphate and reduction of pyruvate to lactate, respectively. To increase the flux through MEP pathway genes, *dxs, dxr, ispD, ispE, ispF, ispG*, and *ispH* are overexpressed. Kumaran et al. (2019) balanced the expression of *ispG* and *ispH* to prevent HMBPP accumulation, which can reduce cell viability and inhibit MEP pathway flux. To reduce the accumulation of MEcPP, genes *ydbK* (pyruvate flavodoxin reductase), *fldA* (flavodoxin), and *fdx* (reduced ferredoxin) are overexpressed. We started the project with the EGN1 strain. The EGN1 strain used valencene synthase to convert farnesene pyrophosphate to valencene and valencene oxidase to further hydroxylate

Table 3.1: Composition of LB medium for WCB preparation.

| Components | Concentration (g/L) |
|---|---|
| Tryptone | 10 |
| Yeast Extract | 5 |
| NaCl | 10 |

valencene and oxidize it to nootkatone. Findings from the process development work for EGN1 were implemented for the AMD strain, which was designed for AMD4,11 production via the heterologous expression of the ADS gene derived from *A. annua*. Lessons from the AMD4,11 production work were tested with the EGN2 strain to further consolidate the results. The EGN2 strain is an improved form of EGN1 with higher MEP flux due to the improved expression of gene encoding along with additional copies of *ispH* and *ispG* reducing the accumulation of MEcPP. For the purpose of confidentiality, genotypic information cannot be disclosed in this thesis.

## 3.2 Media

### 3.2.1 Lysogeny broth medium

Lysogeny broth (LB) medium was used for the preparation of working cell bank (WCB) stock. For this preparation, 25 g of LB broth, Miller powder (BD Difco™) was added to deionized (DI) water to make 1L of solution. The final composition of the medium is specified in Table 3.1. The solution was well mixed before sterilization for 20 min at 121 °C, followed by storage at room temperature.

### 3.2.2 Preculture medium for shake flask (FM-seed)

Preculture cultivations in shake flasks were performed in FM-seed medium. The compositions of the basal salts for FM-seed medium are outlined in Table 3.2. While preparing 1L of medium, all the components were added in 600 g of DI water and pH was adjusted to 7 using 6N NaOH. This was topped up with DI water to final weight of 964.94 g before sterilization for 20 min at 121 °C, followed by storage at 4 °C.

Table 3.2: Medium composition for preculture cultivation for a 1L volume.

| Components | Concentration (g) |
|---|---|
| $(NH_4)_2SO_4$ | 3.6 |
| $KH_2PO_4$ | 12 |
| $K_2HPO_4$ | 4.58 |
| Citric Acid | 1 |
| Yeast extract | 5 |
| Antifoam, Struktol®J647 | 0.5 |

Table 3.3: Post-sterile additions to preculture shake flask cultivations.

| Components | Stock concentration (g/L) | Concentration (ml/L) |
|---|---|---|
| Glucose | 500 | 30 |
| $FeSO_4.7H_2O$ | 27.8 | 1.44 |
| Trace elements (1000x) | NA | 1 |
| $MgSO_4$ | 120.4 | 2.49 |
| Thiamine HCl | 33.7 | 0.13 |

Thereafter, it was supplemented with the solutions mentioned in Table 3.3 prior to inoculation. The amounts were adjusted based on the recipe volume in the shake flask.

### 3.2.3 Preculture medium for bioreactor

While performing scale-up studies, cell suspensions from shake flasks were transferred to a 10L bioreactor to generate a seed train for inoculation into the 300L bioreactor. The compositions of basal salts for cultivation in a 10L bioreactor were the same as for the preculture medium in shake flask cultivation (see Table 3.2). All of the components were added to 6000 g of DI water for a final weight of 6919.51 g before sterilization for 20 min at 121 °C with the bioreactor in situ. Thereafter, it was supplemented with the solutions mentioned in Table 3.3 according to the required volume prior to inoculation. The pH of the media was adjusted to 7 using 6N NaOH.

### 3.2.4 Bioreactor medium (FM)

The starting volume for the bioreactor cultivation was 1L (110L in the case of 300L cultivation). To prepare 1L of FM medium, the basal components listed in Table 3.4

Table 3.4: Medium composition for bioreactor cultivation for a 1L volume.

| Components | Concentration (g) |
|---|---|
| $(NH_4)_2SO_4$ | 7.2 |
| $KH_2PO_4$ | 12 |
| $K_2HPO_4$ | 4.58 |
| Citric Acid | 1 |
| Yeast extract | 10 |
| Antifoam, Struktol®J647 | 0.5 |

Table 3.5: Post-sterile additions to bioreactor cultivations.

| Components | Stock concentration (g/L) | Concentration (ml/L) |
|---|---|---|
| Glucose | 500 | Variable |
| $FeSO_4.7H_2O$ | 27.8 | 1.44 |
| Trace elements (1000x) | NA | 1 |
| $MgSO_4$ | 120.4 | 2.49 |
| Thiamine HCl | 33.7 | 0.13 |
| Safflower oil | NA | 100 |

were dissolved in DI water and weighed up to a final weight of 820.93 g. The resulting solution was added to the bioreactor vessel and sterilized for 20 min at 121 °C. The sterile solution of supplements mentioned in Table 3.5 was added aseptically after the media had cooled down. The glucose concentration varied for different studies during this thesis. The pH of the media was adjusted to 7 using 6N NaOH solution before inoculation. Safflower oil was sterilized for 20 min at 121 °C before being added to the bioreactor medium.

### 3.2.5 Post-sterile stock solutions

#### Glucose solution, 500 g/L

A density of 500 g/L glucose solution was measured to be 1.2 g/ml. To prepare 1L of 500 g/L glucose stock solution, 550 g of glucose monohydrate was added to 650 g of DI water in a 1L bottle with a magnetic stir bar and mixed to form a slurry. It was then sterilized for 20 min at 121 °C. The solution was stirred using the magnetic stirrer to reach homogeneity and then stored at room temperature.

Table 3.6: Composition of 1000x trace element solution.

| Components | Concentration (g/L) |
|---|---|
| $MnCl_2.4H_2O$ | 10 |
| $ZnSO_4.H_2O$ | 2 |
| $CuCl_2.2H_2O$ | 1 |
| $CaCl_2.2H_2O$ | 40 |
| $H_3BO_3$ | 0.5 |
| $Na_2MoO_4.2H_2O$ | 2 |

**Iron sulfate heptahydrate solution (0.1M), 27.8 g/L**

To prepare 1L of 0.1M $FeSO_4.7H_2O$ solution, 27.8 g of $FeSO_4.7H_2O$ (crystalline) was added to 800 ml of DI water. It was gradually titrated with 98% $H_2SO_4$ until all solids were dissolved and a clear solution was obtained. The solution was filtered aseptically through a 0.2$\mu$m Polytetrafluoroethylene (PTFE) membrane filter and stored at 4 °C.

**Magnesium sulfate (1M), 120.4 g/L**

To prepare 1L of 1M $MgSO_4$ solution, 120.4 g of $MgSO_4$ powder was added to DI water to reach a volume of 1L and mixed well until completely dissolved. It was aseptically filtered through a 0.2 $\mu$m PTFE membrane filter and stored at 4 °C.

**Thiamine hydrochloride (0.1M), 33.7 g/L**

To prepare 1L of 0.1M Thiamine HCl, 33.7 g of thiamine hydrochloride powder was added to DI water to reach a total solution volume of 1L and mixed well until completely dissolved. It was aseptically filtered through a 0.2 $\mu$m PTFE membrane filter and stored at 4 °C.

**Trace element solution (1000x)**

All components of the trace element solution mentioned in Table 3.6 were sequentially dissolved in DI water. 10% HCl was gradually added to the mixture until all components were dissolved and a clear solution was obtained. The solution was filtered aseptically through a 0.2 $\mu$m PTFE membrane filter and stored at 4 °C.

### 3.2.6 Bioreactor feed medium

For fed-batch bioreactor cultivation, multiple solutions were fed as a source of carbon, nitrogen, phosphate, and other minerals to avoid nutrient starvation. The following section outlines the preparation of these solutions.

**Feed-1 solution**

To prepare 1L of feed-1, 550 g of glucose monohydrate was added to 650 g of DI water and sterilized in a feed reservoir for 20 min at 121 °C. Then, 1.6 ml of trace element solution (1000x) and 21 $\mu$L of 0.1M thiamine HCl solution were added aseptically once the feed had cooled down.

**Feed-2 solution**

To prepare 1L of feed-2, 550 g of glucose monohydrate was added to 400 g of DI water and sterilized in a feed reservoir for 20 min at 121 °C. Another solution of DI water with 31.28 g of $KH_2PO_4$, 24.36 g of $K_2HPO_4$, and 194.36 g was sterilized for 20 min at 121 °C and added to the feed reservoir aseptically along with 1.6 ml of trace element solution (1000x) and 21 $\mu$L of 0.1M thiamine HCl solution once it had cooled down.

**Feed-3 solution**

To prepare 1L of feed-3, 550 g of glucose monohydrate was added to 300 g of DI water and sterilized in a feed reservoir for 20 min at 121 °C. Another solution with 31.28 g of $KH_2PO_4$, 24.36 g of $K_2HPO_4$, 100 g of YE, and 194.36 g of DI water was sterilized for 20 min at 121 °C and added to the feed reservoir aseptically along with 1.6 ml of trace element solution (1000x) and 21 $\mu$L of 0.1M thiamine HCl solution once it had cooled down.

**Feed-4 solution**

To prepare 1L of feed-4, 660 g of glucose monohydrate, 39.25 g $KH_2PO_4$, and 31.25 g of $K_2HPO_4$ was added to 540 g of DI water at 60 °C. Once dissolved, the feed

solution was filter sterilized through a 0.2 $\mu$m PTFE-cartridge filter assembly into the feed reservoir. Then, 1.92 ml of trace element (1000x) solution and 25 $\mu$L of 0.1 M thiamine HCl solution were added to the feed reservoir through a sterile connection.

**Feed-5 solution**

To prepare 1L of feed-4, 660 g of glucose monohydrate was added to 540 g of DI water at 60 °C. Once dissolved, the feed solution was filter sterilized through a 0.2 $\mu$m PTFE-cartridge filter assembly into the feed reservoir. Then, 1.92 ml of trace element (1000x) solution and 25 $\mu$L of 0.1 M thiamine HCl solution were added to the feed reservoir through a sterile connection.

**Ammonium feed solution**

To prepare 1L of ammonium feed solution, 99 g of $(NH_4)_2SO_4$ was added to 882 g of DI water in a feed reservoir and sterilized for 20 min at 121 °C. Once the reservoir had cooled down, 18.75 ml of 25% $NH_4OH$ was added to it aseptically.

Table 3.7: Devices utilized for the present study.

| **Device** | **Name** | **Supplier** |
|---|---|---|
| Bioreactor and modules | DASGIP® Parallel Bioreactor System-2L, 8-fold system | Eppendorf, Hamburg, Germany |
| | 300L bioreactor | Bioengineering, Wald, Switzerland |
| | Glass bioreactor-10L | Sartorius Stedim Biotech, Göttingen, Germany |
| Bench-top rotary shaker | HAT CH-4204 | Infors HT, Bottmingen, Switzerland |
| Pressure probe-300L | PR-35 X HT KELLER | Druckmesstechnik, Jestetten, Germany |
| Temperature probe-300L/10L | PT100 | Bioengineering, Wald, Switzerland |
| Dissolved oxygen probe-300L/10L | InPro 6800 | Mettler-Toledo, Gießen, Germany |
| pH probe-300L/10L | 405-DPAS-SC-K8S | Mettler-Toledo, Gießen, Germany |
| Exhaust gas analyzer-300L | BCPO2 for oxygen, BCPCO2 for carbon dioxide | BlueSens, Herten, Germany |
| Mass flow controller-300L/10L | MFC GFC171S | Analyt-MTC GmbH, Müllheim, Germany |
| Peristaltic pump | 120 U/DV (200RPM) | Watson Marlow, Falmouth, UK |
| Spectrophotometer | GeneQuant1300 | GE Healthcare Life Sciences, Chicago, USA |
| | DR3900 | Hach, Loveland, USA |
| | DR2800 | Hach, Loveland, USA |
| Centrifuge | 5403 R | Eppendorf, Hamburg, Germany |
| HPLC | Agilent 1200 | Agilent Technologies, Santa Clara, USA |
| HPLC Columns | Rezex ROA organic acid column H+ 8% column 300 × 7.8 mm, 8 $\mu$m | Phenomenex, Torrance, USA |
| | Rezex ROA organic acid column H+ 8% guard column 50 × 7.8 mm, 8 $\mu$m | Phenomenex, Torrance, USA |
| | Supelcosil LC-18-T column 150 × 4.6 mm, 3 $\mu$m | Merck KGaA, Darmstadt, Germany |
| | ZORBAX Eclipse Plus C18 column 250 × 4.6 mm, 5 $\mu$m | Agilent Technologies, Santa Clara, USA |
| | ZORBAX Eclipse Plus C18 guard column 12.5 × 4.6 mm, 5 $\mu$m | Agilent Technologies, Santa Clara, USA |

Table 3.8: Devices utilized for the present study.

| **Device** | **Name** | **Supplier** |
|---|---|---|
| LC-MS QQQ | Agilent 1290 Infinity II LC system with 6460 QQQ | Agilent Technologies, Santa Clara, USA |
| GC | GC 5890 Series II | Agilent Technologies, Santa Clara, USA |
| GC column | Restek 5MS-Sil 0.25mm ID, 0.1 $\mu$m df, 30 m | Restek, Bellefonte, USA |
| Glucose biosensor | LaboTRACE analyzer | TRACE Analytics Gmbh, Braunschweig, Germany |

Table 3.9: Software used for the present study.

| Software | Provider |
|---|---|
| Windows 8 | Microsoft, Redmond, USA |
| Windows Vista | Microsoft, Redmond, USA |
| Office 2013 | Microsoft, Redmond, USA |
| Dasgip® Control | Eppendorf, Hamburg, Germany |
| LABVIEW 2009 SPI | National Instruments, Austin, USA |
| Matlab R2014b | Mathworks, Natick, USA |
| ChemStation | Agilent Technologies, Santa Clara, USA |
| R | The R Foundation, Vienna, Austria |
| DESeq2 | Bioconductor, Boston, USA |

## 3.3 Devices and software

All devices used during the thesis are outlined in Table 3.8. All of the required software are listed in Table3.9.

## 3.4 Seedtrain

Cultivation was performed through various stages of seed train to accumulate biomass for bioreactor cultivation.

### 3.4.1 Cryogenic cultures/Working cell bank

For the preparation of a WCB in the form of cryogenic stock, cells were transferred from the master cell bank provided by Manus Bio Inc. (Cambridge, USA) to a 500 ml conical Erlenmeyer flask containing 100 ml of LB media (see Section 3.2.1). The flask was incubated at 37 °C on a bench-top rotary shaker at 110 RPM for 14–16 h. Then, 500 $\mu$L of cell suspension was mixed well with 500 $\mu$L of 60% glycerol in cryovials. The cryovials were then flash frozen in liquid nitrogen and stored at -70 °C as a WCB.

### 3.4.2 Preculture cultivation

#### Seed culture in shake flasks

For preculture cultivation in shake flasks, 0.8 ml of frozen cryo stock from a WCB

(see Section 3.4.1) was transferred to 400 ml of FM-seed media (see Section 3.2.2) in a 2000 ml non-baffled Erlenmeyer flask. The flask was incubated overnight at 37 °C on a bench-top rotary shaker at 110 RPM and harvested after 12–14 h during the mid-exponential growth phase to be transferred to the next stage of the seed train or production bioreactor.

**Seed culture in bioreactor**

While cultivating in the 300L bioreactor, one more stage of seed train had to be installed in the 10L bioreactor. Then, 8L of FM-seed media (see Section 3.2.2) was prepared for the cultivation, inclusive of 500 ml of cell suspension grown in a 2000 ml shake flask. The reactor was operated at 37 °C, 1.6 sL/min air flow, and constant agitation of 500 RPM. pH was maintained at a setpoint of 7 using 6N NaOH. Cells were cultivated for 12–14 h, followed by transfer to a 300L bioreactor for main cultivation.

## 3.5 Main bioreactor cultivation

Main cultivation was performed in two volumes, as follows.

- 2L Dasgip bioreactor with a starting volume of 1L.
- 300L Bioengineering bioreactor with a starting volume of 110L.

### 3.5.1 2L bioreactor cultivation

Fed-batch cultivations targeting process development for the production of AMD4,11 and EGN were conducted in an 8-fold 2L Dasgip bioreactor. Each vessel was equipped with probes to measure temperature, pH, and dissolved oxygen. Air was sparged into the vessel through a multi-orifice sparger, while the distribution of air bubbles and mixing of liquid was achieved by three Rushton impellers operated with a connected rotor shaft. An external condenser was installed to prevent volume loss due to evaporation. A connection was extended to the exhaust gas analyzer module built with BlueSens technology. The exhaust gas analyzer was calibrated for controlled setpoints

with one level established using air while the other level of the calibration was installed for a gas stream with 2% (v/v) $O_2$ and 10% (v/v) $CO_2$. A pH probe was calibrated externally at 37 °C with estimation of offset and slope using buffers at pH of 4 and 7. The vessel was assembled with probes, a condenser, and a headplate, after which a sterile boundary was established. The bioreactor was sterilized for 20 min at 121 °C with basal media (see Section 3.2.4). Sterile supplements were added via a 60 ml syringe. The $p_{O2}$ probe was calibrated in situ at 37 °C. The probe was calibrated for offset by sparging the medium with pure nitrogen, while the slope was determined using the saturation level of air in the media. The bioreactor was provided with six feeding ports, of which four were connected to feed reservoirs containing 6N NaOH, 6N $NH_4OH$, carbon feed solution (Feed-1/2/3), and ammonium feed solution. The pH was raised to 7 using 6N NaOH. Each vessel was equipped with an individual electrical heating element featuring an integrated safety temperature sensor and separate cooling coils, which can be activated by solenoid valves to maintain the vessels at 37 °C. Figure 3.1 shows the Dasgip reactor system stationed. Process conditions were determined based on the objectives of the study.

### 3.5.2 300L bioreactor cultivation

Scale-up of the process for EGN production was performed in a 300L Bioengineering fermenter equipped with probes to measure temperature, pH, and dissolved oxygen. The pH probe calibration was accomplished as mentioned in Section 3.5.1. The $p_{O2}$ probes were calibrated externally (100% air saturation and 0% oxygen depletion using 1 g of sodium sulfite in 1L of DI water). Air was sparged into the vessel through a multi-orifice sparger, while bubble breakage and distribution along with liquid mixing were achieved using two Rushton impellers. A moisture trap was mounted in line with the vessel to dehumidify the air going toward the BlueSens exhaust gas analyzer. The exhaust gas analyzer was calibrated as per Section 3.5.1. The vessel came with an external jacket meant to supply steam and cooling water. After inserting the probes as well as feed and sample ports, a sterile boundary was established and the bioreactor

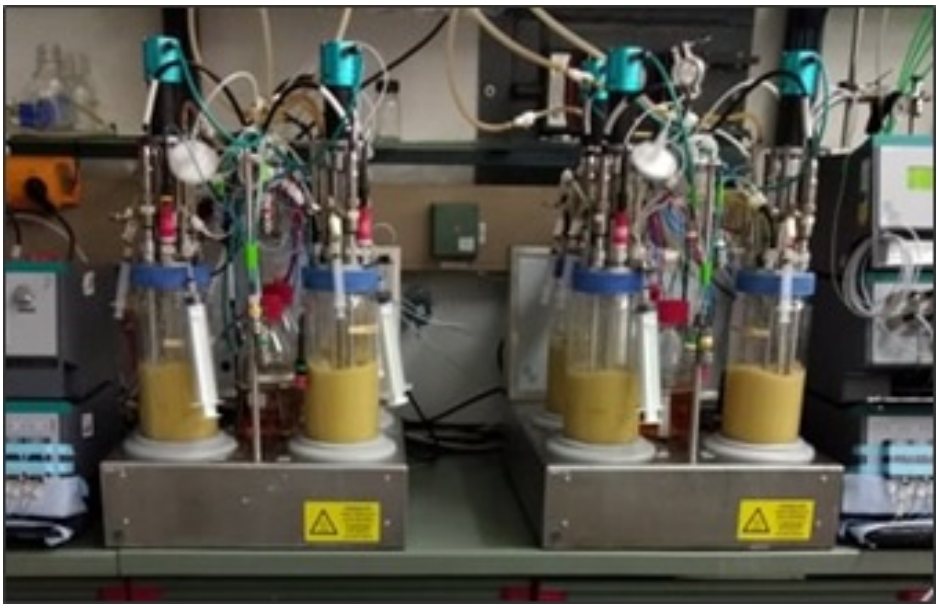


Figure 3.1: Dasgip bioreactors used for the present study.

was sterilized for 20 min at 121 °C with basal media (see Section 3.2.4). Feed ports were connected to 6N NaOH, 6N $NH_4OH$, carbon feed solution (Feed-4/5), and ammonium feed solution. After the addition of sterile supplements, the pH was adjusted to 7 using 6N NaOH. Process conditions were determined using the scale-up criterion.

### 3.5.3 Process control

Soon after the inoculation, $p_{O2}$ began to decrease, indicating an increase in oxygen uptake rate due to respiration. Dissolved oxygen levels were controlled at 40% of air saturation via an agitation cascade designed according to the demands of the study. As soon as the initially supplemented glucose was exhausted, there was a break in metabolism and the oxygen uptake rate steeply decreased, resulting in an $p_{O2}$ spike. This marked the end of the batch phase and carbon feeding was installed in the form of feed-1/2/3/4/5. A carbon limitation strategy was implemented at a 2L scale to avoid byproduct formation. The initial feed rate of glucose was calculated by subjecting the culture to a bolus addition of 2 ml of carbon feed and determining the rate of consumption based on the amount of time required to induce a spike in $p_{O2}$ after exhaustion of the bolus amount. The carbon feed rate was manipulated via the offline monitoring of glucose concentration during the process. However, an exponential

feeding strategy was in place at the 300L scale until 24 h after inoculation, followed by a carbon limitation feeding scheme similar to that of the 2L scale. Throughout the entire process, pH was controlled at 7 using 6N $NH_4OH$. It also acted as a source of nitrogen alongside $(NH_4)_2SO_4$ present in the batch media. As soon as 180 mmol of ammonia were added via $NH_4OH$, the base was switched to 6N NaOH to avoid excess addition of ammonia and exercise control over ammonia concentration in the bioreactor using ammonium feed. To avoid foaming during the process, and antifoam (Struktol J®647) was added at regular intervals. Samples were withdrawn during the process to monitor optical density and glucose concentration and to periodically perform other offline measurements.

### 3.5.4 Sampling

A flow chart of the sampling and its further process is presented in Figure 3.2. The sample port was first flushed with 3–4 ml of cell suspension. Then, 4 ml of the probe was harvested through the port for further processing and divided into three parts. Then, 1 ml of sample was processed for optical density measurement. The supernatant was collected after centrifugation and used for the offline measurement of glucose, ammonia, organic acid, and amino acids. The second part of the sample was used for ethyl acetate extraction of the product (AMD4,11/EGN). One parallel step was added to this sequence when intracellular metabolite analysis was required. For transcriptomic analysis, 1 ml of the sample was withdrawn from the bioreactor and flash frozen in liquid nitrogen to form pellets that could be sent for RNA sequencing. A detailed description of the analysis is outlined in the following section.

## 3.6 Analytical methods

Cultivation was accompanied by the measurements:

- Optical density and dry cell weight
- Substrate concentration using enzymatic assays

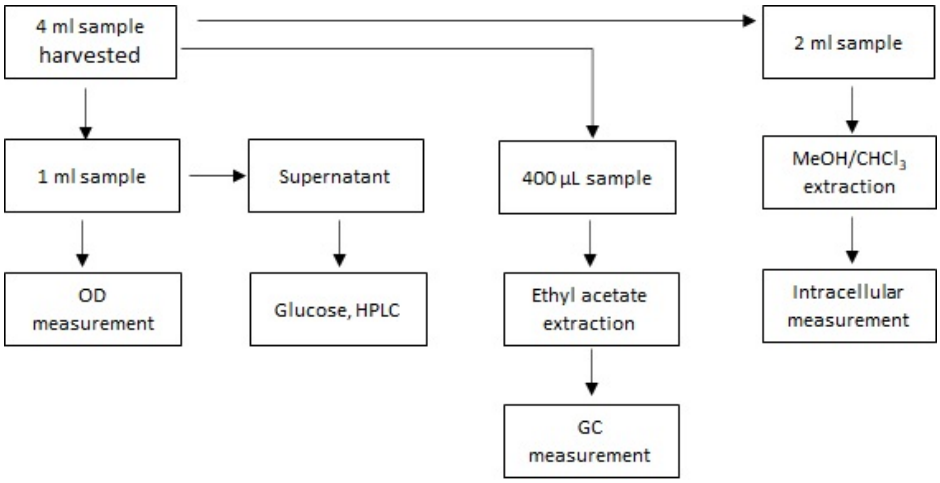



Figure 3.2: Flowchart of sampling during cultivation.

- Organic acids and amino acids using HPLC-DAD/RID/FLD
- Product using GC-FID
- Intracellular metabolite concentration using LC-MS and HPLC-DAD

A detailed protocol for the analysis is described as follows.

### 3.6.1 Optical density and dry cell weight measurement

Optical density and dry cell weight were designated to be the prime attributes used to determine biomass growth. Optical density throughout the experiment was measured as described in Patil et al. (2020). The optical density/biomass correlations for the AMD, EGN1, and EGN2 strains were 0.224, 0.3, and 0.3, respectively.

### 3.6.2 Glucose measurement

To achieve carbon limitation, glucose was measured regularly during the process according to the protocol outlined in Patil et al. (2020).

### 3.6.3 Organic acid measurement

Samples were withdrawn at regular intervals during the experiment. To separate the supernatant, 1 ml of well-mixed broth was centrifuged at 20,173×g and 4°C for

10 min. The aqueous layer of the supernatant was carefully pipetted out and then filtered through ROTILABO® PVDF 0.22 $\mu$m 13 mm syringe filters and stored at -70 °C for further analysis. Acetate and lactate were measured using an HPLC Agilent 1200 with a Rezex ROA organic acid column $H^+$ (8%) column (300 × 7.8 mm, 8 $\mu$m, Phenomenex). Since phosphate interfered with the elution of organic acid during the analysis, a protocol for phosphate removal was implemented for its precipitation from all the samples. To accomplish this, 45 $\mu$L 4M $NH_3$ and 100 $\mu$L 1.2M $MgSO_4$ were added to 1000 $\mu$L of sample. After 5 min of incubation, it was centrifuged for 5 min at 20,173×g and room temperature. Then, 500 $\mu$L of supernatant was transferred to 500 $\mu$L of 0.1M $H_2SO_4$. After thorough mixing and 15 min of incubation at room temperature, the obtained solution was centrifuged for 15 min at 20,173×g at room temperature (RT). Finally, the supernatant was loaded for HPLC injection (10 $\mu$L injection volume). Separation was performed under isocratic conditions at 50 °C (column temperature) for 45 min with 14mM $H_2SO_4$ as mobile phase at a constant flow rate of 0.6 mL/min. Acetate was detected via an Agilent 1200 Series refractive index detector at 32 °C. A six-point calibration curve was used as the external reference standard. Standard solutions were also processed for phosphate precipitation to remove any bias from the calibration curve.

### 3.6.4 Amino acid measurement

The supernatant obtained after filtration of the biosuspension sample was analyzed for its extracellular amino acid concentration. The amino acid concentration was determined using an HPLC Agilent 1200 series apparatus arrayed with an Agilent ZORBAX Eclipse Plus C18 column (250×4.6mm, 5$\mu$m) and fluorescence detector. Pre-column derivatization of the eluent was carried out with ortho-phthaldialdehyde before detection (Lindroth & Mopper, 1979). Notably, this method was adapted from Buchholz et al., 2014. The elution buffer consisted of a polar phase (10mM $Na_2HPO_4$, 10mM $Na_2B_4O_7$, 0.5 mM $NaN_3$, pH 8.2) and a non-polar phase (45% (v/v) acetonitrile, 45% (v/v) methanol). A six-point calibration curve was used as the external reference

standard.

### 3.6.5 Ammonia measurement

The supernatant obtained by spinning down the cell suspension was diluted appropriately and analyzed using LCK302 enzymatic assays provided by Hach Lange. For this process, 0.2 ml of sample was added to the glass cuvette of a LCK302 kit and mixed with the supplied reagents. After incubating the assay for 15 min at room temperature, the ammonia concentration was determined using a DR3900 spectrophotometer (Hach).

### 3.6.6 Phosphate measurement

For phosphate measurement, 0.5 ml of diluted sample was pipetted to the cuvette supplied with an LCK348 enzymatic assay kit and mixed with the provided reagent. After a reaction time of 10 min, phosphate was quantified using a DR2800 spectrophotometer (Hach).

### 3.6.7 Quantification of AMD4,11 and EGN

The extraction of amorpha-4,11-diene was done by treating 400 $\mu$L of biosuspension according to the extraction procedure described in Patil et al. (2020). EGN extraction was also accomplished using the same method. Amorpha-4,11-diene quantification was performed on GC-FID according to the method outlined in Patil et al. (2020).
EGN quantification of the samples was performed on an Agilent 5840 GC equipped with a flame ionization detector using a Restek 5MS-Sil 0.25mm ID column (0.1 $\mu$m df, 30 m length column length) for separation. For this process, 2 $\mu$L of the sample was injected through the injection port, which was maintained at 230 °C. Oven temperature was maintained at 120 °C for 0.5 min and then increased to 180 °C at 15 °C/min. After analysis, the oven was heated to 300 °C and held for 3.5 min to remove any impurities. A five-point calibration curve was developed as the external reference standard.

### 3.6.8 Exhaust gas analysis

Dasgip bioreactors were attached with in line monitoring of exhaust gas composition using a GA4 4-channel exhaust gas analyzer (Dasgip® GA4) equipped with BlueSens technology. The outlet stream through the fermenter was analyzed for the volumetric fraction of oxygen and carbon dioxide using non-dispersive infrared sensors (Eppendorf, Hamburg, Germany). The 300L bioengineering vessel was assembled with an inline single-channel photometric infrared gas analyzer to facilitate the estimation of volumetric fraction of oxygen and carbon dioxide (BPO2 and BPCO2, BlueSens, Herten, Germany).

### 3.6.9 Intracellular metabolite measurement

The extraction and quantification of intracellular ATP and MEP intermediates were performed as explained in Patil et al. (2020). The analysis and quantification were outsourced to Manus Bio Inc. (Cambridge, USA).

## 3.7 Data processing and analysis

For a comprehensive understanding of the process and exploration of the relationship between different parameters, derived kinetic variables were calculated as explained in the following section. In the present study, a fed-batch process was conducted. Hence, the mass balance was applied to the open system.

### 3.7.1 Growth rate

The specific growth rate of the system $\mu$ (in $h^{-1}$) was derived from the mass balance of biomass for the open system and $\lim \triangle t \rightarrow 0$ as:

$$\frac{dm_x}{dt} = \mu c_x V_r \tag{3.1}$$

where $m_x$ is the total biomass (in g), $c_x$ is the concentration of biomass (in g/L), and $V_r$ is the reaction volume (in L). Simultaneously, $m_x = c_x V_r$. Hence, Equation (3.1) can be written as:

$$\frac{dm_x}{dt} = \frac{d(c_x V_r)}{dt} \tag{3.2}$$

Thus, from Equations (3.1) and (3.2),

$$\mu dt = \frac{d(c_x V_r)}{c_x V_r} \tag{3.3}$$

Integrating Equation (3.3) over a period of $t_0$ to $t_1$, we obtain

$$\mu = \ln(\frac{c_{x,1} V_{r,1}}{c_{x,0} V_{r,0}}) \frac{1}{t_1 - t_0} \tag{3.4}$$

In the above Equation (3.4), $c_{x,1}$ and $V_{r,1}$ are biomass concentration and reaction volume at time $t_1$ (in h), respectively, while $c_{x,0}$ and $V_{r,0}$ are biomass concentration and reaction volume at time $t_0$ (in h), respectively. Equation (3.4) was used to calculate the specific growth rate of biomass over a given time interval during the fed-batch process.

### 3.7.2 Specific substrate consumption rate

The specific substrate consumption rate $q_s$ (in g/g/h) of an *E. coli* strain in the fed-batch process is derived from the mass balance on substrate:

$$\frac{dm_s}{dt} = c_F F_{in} - q_s c_x V_r \tag{3.5}$$

where $m_s$ is mass of the substrate (in g) and $F_{in}$ is the feed rate of substrate entering the bioreactor (in L/h) with a concentration of $c_F$ (in g/L). Now, $m_s = c_s V_r$. Therefore, we have:

$$\frac{dm_s}{dt} = \frac{d(c_s V_r)}{dt} \tag{3.6}$$

where $c_s$ is the substrate concentration in the bioreactor (in g/L). From Equation (3.6) in (3.5), we obtain:

$$\frac{d(c_s V_r)}{dt} = c_F F_{in} - q_s c_x V_r \tag{3.7}$$

Rearranging Equation (3.7) and integrating over a time period of $t_0$ to $t_1$, we obtain:

$$\int (c_F F_{in})dt - (c_{s,1} V_{r,1} - c_{s,0} V_{r,0}) = q_s \int (c_x V_r)dt \tag{3.8}$$

where $c_{s,1}$ the is substrate concentration at time $t_1$ and $c_{s,0}$ is the substrate concentration at time $t_0$. For a constant feed rate $F_{in}$ and substrate feed concentration $c_F$, $\int c_F F_{in} dt = c_F F_{in}(t_1 - t_0)$. Furthermore, from Equation (3.1), we obtain:

$$c_F F_{in}(t_1 - t_0) - (c_{s,1} V_{r,1} - c_{s,0} V_{r,0}) = q_s (c_{x,1} V_{r,1} - c_{x,0} V_{r,0}) \frac{1}{\mu} \tag{3.9}$$

By rearranging Equation (3.9), we obtain:

$$q_s = \mu \frac{c_F F_{in}(t_1 - t_0) - (c_{s,1} V_{r,1} - c_{s,0} V_{r,0})}{(c_{x,1} V_{r,1} - c_{x,0} V_{r,0})} \tag{3.10}$$

Equation (3.10) was used for the calculation of the biomass-specific substrate rates for glucose, phosphate, and amino acids during the thesis.

### 3.7.3 Specific production rate

The specific production rate $q_p$ (in g/g/h) for AMD4,11 and EGN in the fed-batch process is derived from the mass balance of product.

$$\frac{dm_p}{dt} = q_p c_x V_r \tag{3.11}$$

where $m_p$ is mass of product (in g) and now $m_p = c_p V_r$. Therefore, we have:

$$\frac{dm_p}{dt} = \frac{d(c_p V_r)}{dt} \tag{3.12}$$

where $c_p$ is the product concentration in the bioreactor (in g/L). From Equation (3.11) in (3.12), we obtain:

$$\frac{d(c_p V_r)}{dt} = q_p c_x V_r \tag{3.13}$$

Rearranging Equation (3.13) and integrating it over a time period of $t_0$ to $t_1$, we obtain:

$$(c_{p,1} V_{r,1} - c_{p,0} V_{r,0}) = q_p \int (c_x V_r) dt \tag{3.14}$$

where $c_{p,1}$ is the product concentration at time $t_1$ and $c_{p,0}$ is the product concentration at time $t_0$. Similar to the derivation of $q_s$ from Equation (3.1), we obtain:

$$q_p = \mu \frac{(c_{p,1} V_{r,1} - c_{p,0} V_{r,0})}{(c_{x,1} V_{r,1} - c_{x,0} V_{r,0})} \tag{3.15}$$

Equation (3.15) was used to estimate the specific productivity of the process for AMD4,11 and EGN for this thesis.

### 3.7.4 Oxygen consumption rate

As the air was sparged into the bioreactors, cells used the oxygen present in air for the oxidation of substrate via respiration. The volumetric oxygen demand ($Q_{O2}$) of the cell was assumed to be equal to the Oxygen Uptake Rate ($OUR$) of the cell (in mol/L/h). For the calculation of oxygen uptake rate, it was assumed that oxygen concentration in the gas phase was in equilibrium with the dissolved oxygen concentration in the liquid phase. We obtain:

$$OUR.V_r = \triangle \dot{n}_{O2} \tag{3.16}$$

where $\triangle \dot{n}_{O2}$ (in mol/h) is the difference in the inflow and outflow of molar flux of oxygen, $\triangle \dot{n_{O2}} = \dot{n}_{O2,in} - \dot{n}_{O2,out}$, $\dot{n}_{O2,in}$ is the molar flow rate of oxygen coming into the reactor, and $\dot{n}_{O2,out}$ is the molar flow rate of oxygen going out of the reactor. Assuming the ideal gas law, $PV = nRT$ to air. By applying it to the molar flux, we obtain:

$$\triangle \dot{n}_{O2} = (\dot{V}_{gas,in} y_{O2,in} - \dot{V}_{gas,out} y_{O2,out}) \frac{p}{RT} \tag{3.17}$$

where $\dot{V}_{gas,in}$ and $\dot{V}_{gas,out}$ are the volumetric gas flow rate of the inlet gas stream and outlet gas stream (in L/h), respectively. $y_{O2,in}$ and $y_{O2,out}$ is the volumetric fraction of oxygen in inlet and outlet gas stream, respectively. $p$ and $T$ are the pressure and temperature of the mass flow controller for the exhaust gas analyzer. Since nitrogen is the inert gas during cultivation, we can calculate the outlet gas flow rate using the following equation:

$$\dot{V}_{gas,out} = \dot{V}_{gas,in} \frac{(1 - y_{O2,out} - y_{CO2,out})}{(1 - y_{O2,in} - y_{CO2,in})} \tag{3.18}$$

Consequently,

$$OUR = \frac{p\dot{V}_{gas,in}}{RTV_r} [y_{O2,in} - y_{O2,out} \frac{(1 - y_{O2,out} - y_{CO2,out})}{(1 - y_{O2,in} - y_{CO2,in})}] \tag{3.19}$$

The exhaust gas analyzer enables the measurement of the volumetric fraction of oxygen and carbon dioxide in the off-gas stream. This facilitated the derivation of the oxygen uptake rates of cells at a given time during the process. For the calculation of the biomass-specific oxygen uptake rate $q_{O2}$ (in mol/g/h), the following equation could be used:

$$q_{O2} = \frac{OUR}{c_x} \tag{3.20}$$

Most calculations were performed over a time interval between samplings. $q_{O2}$ estimation over a time interval $(t_1 - t_0)$ was calculated using the following equation:

$$q_{O2,(t1-t0)/2} = OUR_{(t1-t0)/2} V_{r,(t1-t0)} \frac{2}{c_{x,t1} V_{r,t1} + c_{x,t0} V_{r,t0}} \tag{3.21}$$

where $q_{O2,(t1-t0)/2}$ is the specific oxygen uptake rate at time $(t_1 - t_0)/2$ and $OUR_{(t1-t0)/2}$ corresponds to the volumetric oxygen uptake rate calculated from the output of the exhaust gas analyzer at time $(t_1 - t_0)/2$.

### 3.7.5 Carbon dioxide evolution rate

Considering parallelism with the $OUR$ calculations, the carbon dioxide evolution rate (in mol/L/h) was determined by imposing a molar balance of carbon dioxide and considering an open system. The following equation was used during the present study:

$$CER = \frac{p\dot{V}_{gas,in}}{RTV_r}[y_{CO2,out}\frac{(1 - y_{O2,out} - y_{CO2,out})}{(1 - y_{O2,in} - y_{CO2,in})} - y_{CO2,in}] \tag{3.22}$$

Dividing $CER$ by the total biomass concentration in the bioreactor yields the biomass-specific carbon dioxide evolution rate (in mol/g/h), which is calculated as follows:

$$q_{CO2} = \frac{CER}{c_x} \tag{3.23}$$

Similar to $q_{O2}$, most calculations were performed over a time interval between samplings. $q_{CO2}$ estimation over a time interval $(t_1 - t_0)$ was calculated using the following equation:

$$q_{CO2,(t1-t0)/2} = CER_{(t1-t0)/2}V_{r,(t1-t0)}\frac{2}{c_{x,t1}V_{r,t1} + c_{x,t0}V_{r,t0}} \tag{3.24}$$

where $q_{CO2,(t1-t0)/2}$ is the specific carbon dioxide evolution rate at time $(t_1 - t_0)/2$ and $CER_{(t1-t0)/2}$ corresponds to the volumetric carbon dioxide evolution rate calculated from the output of the exhaust gas analyzer at time $(t_1 - t_0)/2$.

### 3.7.6 Respiratory quotient

The Respiratory Quotient (RQ) can be translated as a snapshot of the metabolic state of *E. coli* cells. Substrate usage could also be gauged using RQ. Due to the presence of YE in the batch media, the strain would undergo the preferential consumption of carbon sources during cultivation. This was indicated by RQ, which is calculated as:

$$RQ = \frac{CER}{OUR} \tag{3.25}$$

### 3.7.7 Yield of biomass on substrate

The yield of biomass on substrate $Y_{xs}$ (in g/g) is the ratio of the specific growth rate and consumption rate of substrate:

$$Y_{xs} = \mu / q_s \tag{3.26}$$

Thus, from Equation (3.10), we obtain:

$$Y_{xs} = \frac{(c_{x,1}V_{r,1} - c_{x,0}V_{r,0})}{c_F F_{in}(t_1 - t_0) - (c_{s,1}V_{r,1} - c_{s,0}V_{r,0})} \tag{3.27}$$

Equation (3.27) was used for the calculation of yield of biomass on glucose and amino acids for this thesis. Simultaneously, the composition of biomass pertaining to a particular substrate $Y_{sx}$ (in g/g) is defined as:

$$Y_{sx} = \frac{1}{Y_{xs}} \tag{3.28}$$

Thus, to calculate $Y_{sx}$ for amino acids, we used:

$$Y_{sx} = \frac{c_F F_{in}(t_1 - t_0) - (c_{s,1}V_{r,1} - c_{s,0}V_{r,0})}{(c_{x,1}V_{r,1} - c_{x,0}V_{r,0})} \tag{3.29}$$

Values for $Y_{xs}$ and $Y_{sx}$ are reported in molar units in some sections. When calculating them, the molecular weight of the corresponding substrate can be used to modify equations (3.27) and (3.29).

## 3.8 Flux balance analysis

In this thesis, we used the COBRA (Constraint-Based Reconstruction and Analysis) toolbox of MATLAB for the quantitative prediction of cellular flux based on experimental constraints. iJR904 was used as the base *E. coli* model and further modifications were implemented according to strain construct. The curation of the model is described in Patil et al. (2020). The ATP synthesis reaction was deputed as the objec-

tive function to determine the reciprocity of specific glucose and oxygen uptake rates. The protocols described in Schellenberger et al. (2011) were used to compute flux distribution. The net formation/consumption rate of a specified metabolite was calculated by the summation of fluxes for all the reactions incorporating it. This analysis was evolved to draft a window of optimum describing the limits of specific substrate consumption rate for peak production. We also used the base model for assessing the impact of amino acid supplementation on specific growth rate and production of AMD4,11. Fluxes were calculated for two instances to optimize specific growth rate and AMD4,11 productivity independently for the individual uptake of specific amino acids. We confined the study to 17 amino acids and compared it against the control condition (glucose uptake only). To lay out the window of optimum, we used the function "doublerobustnessanalysis" to simulate the flux distribution of the stoichiometric model over a wide range of glucose and oxygen uptake rates with the aim of optimizing AMD4,11 production.

## 3.9 Scale-up criterion

In this thesis, we worked on scaling up the process from a 2L to a 300L bioreactor. Successful scale-up was characterized by replicating the specific productivity of the process across the volume range. For a given scale, some of the criteria described below were calculated and targeted to match while traversing through the volume increase along different scales of bioreactors.

**Constant volumetric aeration rate (VVM)**

The flow rate of air sparged into the bioreactor per unit reaction volume is given as:

$$\dot{V}_{300L} = V_{r,300L} \frac{\dot{V}_{2L}}{V_{r,2L}} \tag{3.30}$$

where $\dot{V}_{300L}$ is aeration rate for the 300L scale (in L/min), $\dot{V}_{2L}$ is aeration rate for the 2L scale (in L/min), $V_{r,300L}$ is the initial reaction volume for 300L, and $V_{r,2L}$ is the

initial reaction volume for 2L.

### Constant tip speed

The scale-up procedure using the constant impeller tip speed criterion (Aiba et al., 1973) can be written as:

$$n_{300L} = n_{2L}\frac{di_{2L}}{di_{300L}} \tag{3.31}$$

In Equation (3.31), $n_{300L}$ and $n_{2L}$ are impeller agitation speeds (in RPM) for the 300L and 2L scales, respectively. $di_{300L}$ and $di_{2L}$ are the impeller diameters (in m) for the 300L and 2L scales, respectively.

### Volumetric power input

Power consumption (input) for the bioreactor to maintain a certain impeller speed is one of the most important variables in bioreactor design. This is defined as (Asenjo, 1994):

$$P = N_p \rho n^3 di^5 N \tag{3.32}$$

$P$ is the ungassed power input (in Watts), $N_p$ is the power number of the impeller, N is the number of impellers, and $\rho$ is the density of the liquid. Volumetric power input can be calculated by dividing the power input with the reaction volume of the bioreactor.

### Volumetric oxygen transfer coefficient ($k_La$)

The empirical estimation of the volumetric oxygen transfer coefficient was calculated using van't Riet's correlation for non-coalescing liquid (Van't Riet, 1979):

$$k_La = \gamma(\frac{P}{V_r})^\alpha (v_g)^\beta \tag{3.33}$$

In the latter equation, the constants are $\gamma$=0.002, $\alpha$=0.7 and $\beta$=0.2.

## 3.10 Transcriptomics

### RNA sequence analysis

Samples harvested from the bioreactor were flash frozen in direct contact with liquid nitrogen. Then, the obtained cell pellets were sent for RNA sequencing to Genewiz, NJ, USA. An Illumina Ribo-Zero rRNA removal kit was used for rRNA depletion. Quality was assessed by an Agilent fragment analyzer. Libraries were sequenced as paired-end reads (150 bp read length) on a HiSeq platform (Illumina) at a depth of 35–44 million reads each.

### Read mapping and gene counting

Trimmomatic v. 0.32 (Bolger et al., 2014) was used to remove adapters and low-quality reads (<Q20) checked by FastQC reports. Genes were aligned to the National Center for Biotechnology Information (NCBI) *E. coli* K12 MG1655 reference genome (RefSeq: NC_000913.3) using the RNA sequencing aligner Bowtie2 v. 2.3.2.2 (Langmead & Salzberg, 2012). On average, the mapping of unique reads covered 77.0% of the total. Aligned reads were counted for each gene based on the corresponding annotation available from the NCBI for the chosen reference sequence by using HTseq-count v. 0.6.1 (Anders et al., 2015) in union mode. On average, 68.4% of the sequenced reads could be assigned uniquely to annotated features. Sequencing depth was approximately 39 million reads per sample, on average. The low mapping percentage is due to three low-quality (RNA Integrity number: RIN  5, Table 3.10, 3.11) samples (i.e., samples referring to one 50.5 h and both 70.75 h replicates of Condition 1, Con1). Reference state for both conditions, Condition 1 and Condition 2 is designated by Ref_9h_1 and Ref_9h_2 respectively. They refer to the samples take at estimated cultivation time of 9h. Low RIN values might be due to RNA degradation in the samples owing to excessive stress. Either the degradation started during fermentation (biologically motivated degradation) or the degradation occurred in isolated

Table 3.10: RIN values-series e-A. Both replicates of Condition 1 70.75 h were excluded from further analysis.

| Samples | Ref_9h_1 | Con1_26h | Con1_50.5h | Con1_70.75h |
|---|---|---|---|---|
| RIN Replicate 1 | 10 | 8.5 | 8.6 | 3.3 |
| RIN Replicate 2 | 9.5 | 7.5 | 4.7 | 5.0 |

Table 3.11: RIN values-B.

| Samples | Ref_9h_2 | Con2_26h | Con2_50.5h | Con2_70.75h |
|---|---|---|---|---|
| RIN Replicate 1 | 9.8 | 10.0 | 9.9 | 9.9 |
| RIN Replicate 2 | 10.0 | 10.0 | 9.9 | 9.7 |

RNA through free RNAses (random degradation). If the degradation is random, the trend of the transcriptomic analysis remains visible and can be addressed via data normalization. Observations from Romero et al. (2014) indicate that useful data can be collected using RNA sequencing (even from highly degraded samples) on the condition that the RIN scores are not associated with the effect of interest in the study. To evaluate the impact of low-quality reads on the dataset, a Pearson correlation was used. Outliers (Pearson correlation coefficient $< 0.3$) were observed in the two biological replicates after 70.75 h of Condition 1 over all samples, meaning that there is no linear correlation. Due to the low RNA quality of both samples at time point 70.75 h (RIN: 3.3 and 4.7 and a Pearson correlation coefficient $< 0.3$), these samples were excluded from the analysis. However, despite the low-quality results of one sample at time point 50.5 h, this sample correlates and clusters well with its corresponding replicate and was thus included in the transcriptomic analysis.

**Transcriptome data analysis**

Differential gene expression analysis was performed using the R package DeSeq2 v. 1.26.0 (Love et al., 2014), available from Bioconductor (Gentleman et al., 2004) (http://www.bioconductor.org). Prior to statistical analysis, all residual non-protein encoding RNA molecules (i.e., tRNA, rRNA, and sRNA) were removed from the HTseq-derived raw count data and a non-specific filter was applied to remove low coverage genes with fewer than two counts per million (54 reads on average) in all

two replicates per condition. All filtering steps caused deviations of less than 5 % from the raw data. Samples were grouped by replicates and an experimental design was chosen that used sample time and treatment (Condition 1, Con1 and Condition 2, Con2) as a combined environmental factor. To normalize the read counts for the comparison of sequencing depth and RNA composition, we used the median of ratios method in DESeq2 to derive a scaling factor. Dividing the original read counts by the scaling factor normalized the generated count values. To model count-based expression data, DESeq2 uses a negative binomial model as a distribution assumption and fits the expression data for each gene to a generalized linear model (GLM). The resulting p-values were adjusted for multiple testing according to Benjamini & Hochberg (1995) to control the false discovery rate (FDR). Genes were identified as significantly differentially expressed by applying FDR-adjusted p-values $< 0.01$ and a $log_2$-fold change $|1|$.

A principal component analysis was used to display the sample-to-sample distances calculated using the DESeq2 package. The principal component analysis was performed using plotPCA.san, which is available on Github (https://gist.github.com/sansense/3399064897f1252d31b23ea5178c033c).

Gene set enrichment and an overrepresentation analysis of up- and downregulated genes were performed using the Bioconductor R-package gage v. 2.36.0 (Luo et al., 2009). Gage tests whether the mean fold change of a gene subset differs significantly from the background based on a one-sample two-tailed t-test. Genes were identified as significantly different with an FDR adjusted p-value $< 0.01$ (Benjamini & Hochberg, 1995). Since tests for gene sets with an extremely small or large number of genes are not robust, only pathways with 10–500 genes were analyzed in GAGE. Functional annotations were derived from the experimental sigma factor-gene interaction dataset from RegulonDB v. 10.6.3 (Salgado et al., 2004) and the Gene Ontology (GO) Groups database with the function go.gsets from gage (Luo et al., 2009). Overall, 3380 out of 4058 genes (83%) could be mapped to GO terms and 3935 out of 4058 genes could be assigned to the sigma factor-gene interaction database from RegulonDB. Further-

more, the Bioconductor package pathview was used to analyze changes in metabolic level (Luo & Brouwer, 2013).

Sample description, mapping statistics, raw read counts, and processed data can be found in the Supplementary Information. Data analysis was performed using the free statistical computing environment R v. 4.0.0 (R Foundation).

# Chapter 4

# Results and Discussion

Process development for the production of terpenoids via the MEP pathway involved five parts. The goal of this thesis was to enhance flux through the MEP pathway, classified by the production of IPP and DMAPP. These two building blocks could be reworked to form either EGN or AMD4,11, and further converted to terpenoids. In this dissertation, plots with dotted lines indicate an approximation of the trend and the actual value between the sample points could lie on either side of the line. Sparsely dotted connection indicates larger deviation from linearity. On the other hand, solid lines indicate the linear trend between the two sample points of the plot. We used arbitrary units to indicate EGN profile. EGN titer was expressed as %AU/L and specific productivity as %mAU/g/h. EGN titer for series A was designated as 100 %AU/L and titer for all the other studies were calculated accordingly.

## 4.1 Challenges in MEP process development manifested by EGN

Process development to produce terpenoids via the MEP pathway was first conducted with EGN1 strain. EGN1 overexpressed the MEP pathway to generate IPP and DMAPP, which was further converted to FPP. FPP was altered to form a terpene precursor of EGN. We experienced no accumulation of FPP. Hence, EGN titer could

be used as an indicator of MEP flux, with 1 %AU of EGN equaling 19.62 mol of IPP. Thus, the conclusions of this work could be inspected against AMD4,11 production.

### 4.1.1 Series A

A fed-batch process in duplicate mode was installed in 2L Dasgip bioreactors. Bioreactor media (see Section 3.2.4) was prepared with a glucose concentration of 20 g/L. 1L of batch media was composed of 40 mg of $FeSO_4.7\,H_2O$, 10 mg $MnCl_2.4\,H_2O$, 3.2 mg of $ZnSO_4.7\,H_2O$, 1mg of $CuCl_2.2\,H_2O$, 40 mg of $CaCl_2.2\,H_2O$, 0.50 mg of $H_3BO_3$, 2 mg of $Na_2MoO_4.2\,H_2O$, 300 mg of $MgSO_4$, 4 mg of thiamine hydrochloride, 7.2g of $(NH_4)_2SO_4$, 12 g of $KH_2PO_4$, 4.58 g of $K_2HPO_4$, and 10 g of YE and 1 g of citric acid. This was a biphasic system with 10% safflower oil (included in the initial volume) added as an organic overlayer to capture EGN produced during fermentation. Process control is described in following section.

#### Process control

The process was operated as mentioned in Section 3.5.3, with a few exceptions. $p_{O2}$ level was controlled at 40% of air saturation via an agitation cascade with a stir speed starting at 400 RPM and capped at 1200 RPM. Aeration was installed at 0.2 vvm, which was calculated based on the initial recipe volume of the vessel. With the exhaustion of initially loaded glucose marked by a spike in $p_{O2}$, feed-1 (see Section 3.2.6) was pumped in to satisfy the carbon demands of cells. Additionally, a carbon limitation strategy was implemented to avoid byproduct formation (Yee & Blanch, 1992). The carbon feed rate was manipulated by the offline monitoring of glucose concentration during the process to maintain a residual glucose concentration close to 0 g/L. During the entire process, pH was maintained at 7 using 6N $NH_4OH$, which was eventually substituted with 6N NaOH to avoid ammonia accumulation. It has been reported that ammonia inhibits the growth of *E. coli* above 170 mM (Thompson et al., 1985). Additionally, ammonia concentrations above 100 mM can inhibit the expression of heterologous proteins (Tsuruta et al., 2009). This process was run for 62 h.

**Process breakdown**

The highest biomass concentration obtained during the cultivation was 34.2±0.36 g/L (Figure 4.1-a). After 35 h, the total biomass in the bioreactor began to decrease, indicating cell death. It should be noted that due to absence of sampling event between inoculation and 35 h, biomass and EGN production trend can deviate from linearity. This is indicated by dotted connections between the two data points. Density of the dots increase after that, indicating closer approximation to linearity. Similar approximation can be applied to all the plots for this study. This was due to eventual nutrient exhaustion leading to a stress response due to starvation, which results in cell lysis (Noor, 2015). This could be correlated to excessive foaming in the bioreactor due to the disintegration of the outer membrane and release of proteins (Shimizu, 2014). Also, the process could not be ideally controlled and there were phases involving glucose starvation and excess that led to stress-related responses in *E. coli*. It has been shown that the circulation of cells around zones of substrate excess and limitation can lead to significant biomass yield losses (Vasilakou et al., 2020). EGN quantification revealed that a maximum titer of 99.92±6.59 %AU/L during cultivation (Figure 4.1-b). Linearity was maintained in the EGN profile until 48 h, after which production apparently ceased.

The biomass-specific productivity of EGN $q_{EGN}$ was 116.78±2.33 %mAU/g/h at 17.5 h, after which it exhibited a gradual depreciation to 47.98±17.60 %mAU/g/h at 55 h. Cells grew at a specific growth rate of 0.17±0.00 $h^{-1}$ at 17.5 h and eventually exhibited a negative value at 41.5 h (Figure 4.2-b). Figure 4.2-a conveys that EGN production is governed by mixed growth kinetics. Biomass-specific productivity depends on the specific growth rate of the cells during the growth, stationary, and post-stationary phases, while production depends on the total biomass present in the bioreactor.

Glucose concentration could not be maintained at near-zero values during most of the fermentation process (Figure 4.3-a). Standard deviation between the two duplicates indicated a highly variable process. However, both the duplicates reached a residual

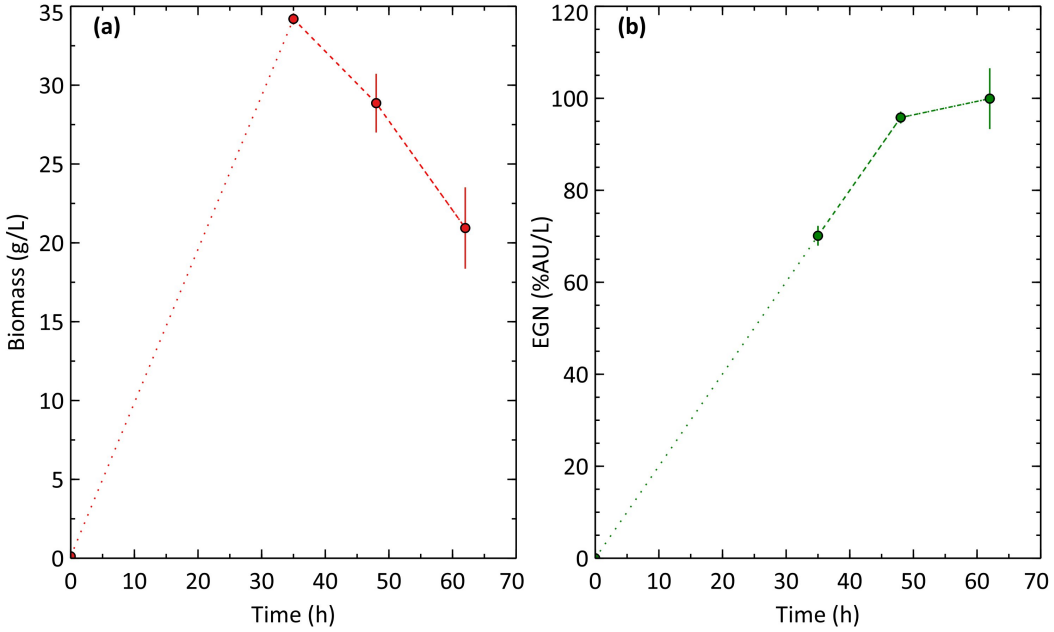



Figure 4.1: Cell density and the production of EGN for Series A. a: Biomass concentration in g/L; b: EGN titer in %AU/L during the time course of cultivation.

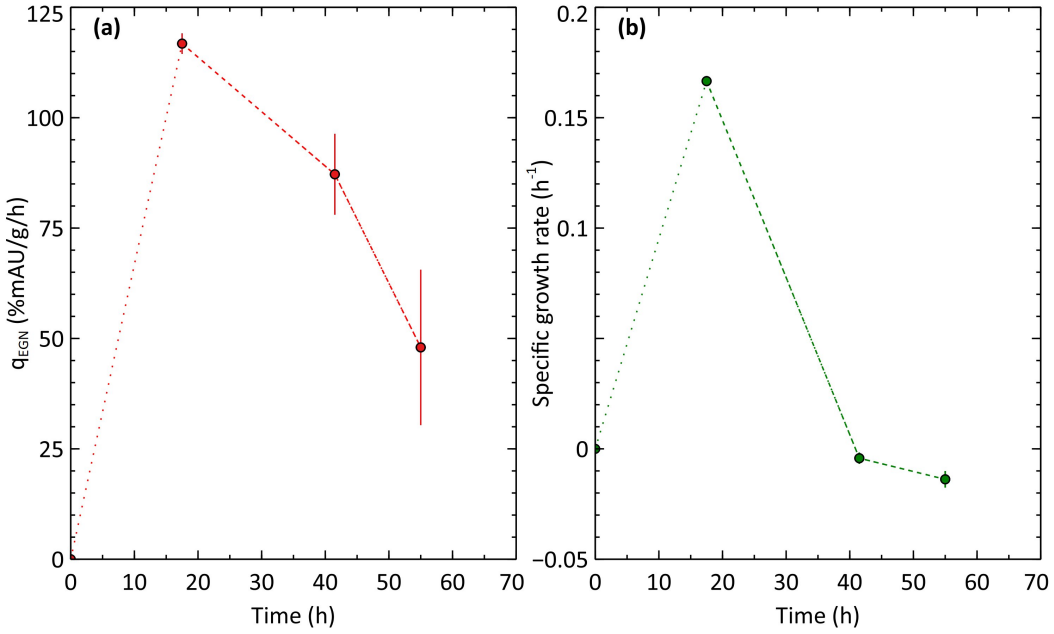



Figure 4.2: Production and growth performance for Series A. a: Biomass-specific productivity, $q_{EGN}$ in %mAU/g/h; b: Specific growth rate for EGN1, $\mu$ in $h^{-1}$.

glucose concentration of zero at the end of the process. Byproduct accumulation in the form of acetate and lactate can also be seen during cultivation. At the end of cultivation, 5.68±0.00 g/L of acetate was accumulated. *E. coli* tends to accumulate acetate when subjected to a fluctuating C:N ratio or if glucose accumulation occurs, which is applicable to our case. However, cells were also enveloped by an oxygen-limited environment for a certain length of time since the $p_{O2}$ levels had dropped to zero. Notably, this scenario could lead to substrate reduction via NADH to form acetate (Eiteman & Altman, 2006). Lactate formation amounted to 9.03±0.00 g/L. Also, lactate and acetate alone accounted for 10.61% of the C-mol fed from glucose. For the *E. coli* wild type strain, a typical yield of biomass on glucose is approximately 0.5 g/g (Shiloach & Fass, 2005). EGN1 could barely reach 0.147 g/g due to the loss of carbon flux to byproducts and the metabolic burden of the overexpressed heterologous pathway (Wu et al., 2016) downstream of MEP. The carbon molar yield of IPP on glucose was 2.03%. *In silico* profiling of the MEP pathway in *E. coli* by Gruchattka et al. (2013) revealed a theoretical yield of IPP on glucose of 0.57 C-mol/C-mol in the case of biomass formation. This highlights a blatant gap between theoretical and empirical performance. However, we could improve our results by addressing some of the obvious issues. At this point, our process was dealing with problems of high variance, a breakdown in productivity midway, and the glaringly low yield on glucose.

### 4.1.2 Phosphate starvation: A malefactor for breakdown

The operating conditions of Series A were scaled up to perform cultivation in a 300L fermenter. FM media (3.2.4) was prepared with an initial glucose concentration of 5 g/L. Glucose amount was trimmed to avoid the formation of byproducts early in the bioprocess due to the heterogeneities in a large scale bioreactor, resulting in the creation of feast-famine zones in the vessel (Garcia-Ochoa & Gomez, 2009). All the other components had the same concentration as Series A. Similar to Series A, a 10%

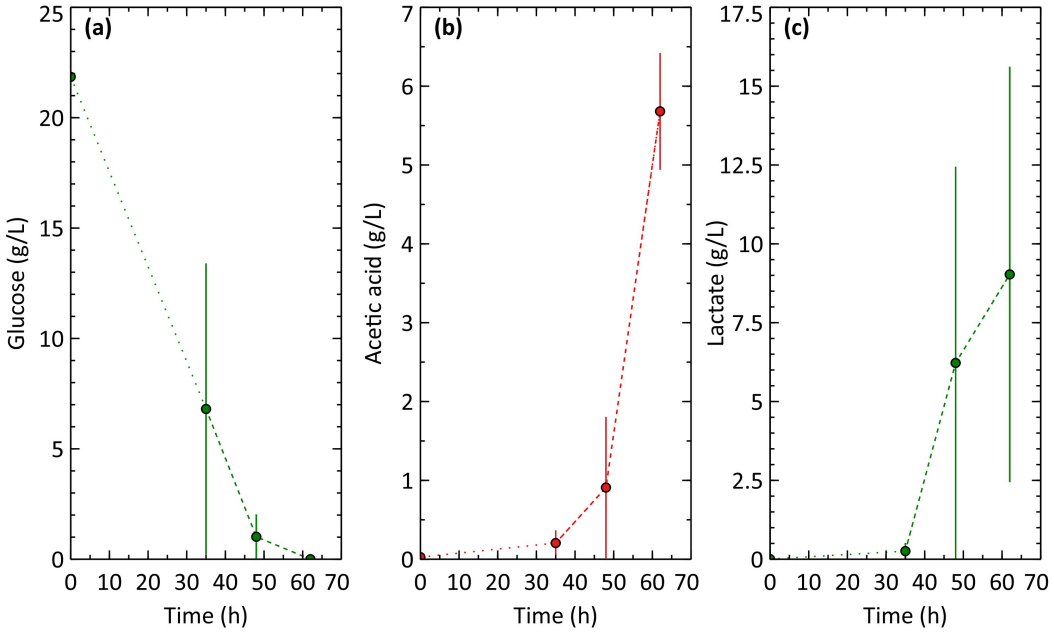



Figure 4.3: Substrate profile and byproduct formation for Series A. a: Glucose concentration in g/L; Center: Acetate concentration in g/L; b: Lactate concentration in g/L.

safflower oil organic overlayer was added to the bioreactor. This study was identified as Series B.

Glucose was monitored closely during the process and, based on the offline measurement exponential feeding strategy, was implemented using feed-5 when the glucose concentration was approximately 1 g/L to avoid starvation. The set point for $p_{O2}$ was lowered to 30% to account for the increased solubility of oxygen due to hydraulic pressure in the bioreactor in addition to a backpressure of 1.2 psig. Dissolved oxygen levels were controlled using agitation speeds starting from 100 RPM and capped at 300 RPM. Based on the constant tip speed criteria for scale-up (see Section 3.9), stir speed was calculated as 276.9 RPM (corresponding 1200 RPM at the 2L scale). To avoid any potential mass transfer issues, we increased the upper limit to 300 RPM. The aeration rate was kept constant throughout the cultivation at 0.2 vvm.

For the exponential feeding strategy, the feed rate was calculated to achieve a biomass-specific growth rate of 0.08 $h^1$ with biomass concentration while starting the feed rate $c_{x,t0}$ at 8 g/L, a yield of biomass on glucose $Y_{xs}$ at 0.9 g/g, and maintenance coefficient

$m_s$ at 0.057 g/g/h. The equation to calculate the feed rate is given as Equation (4.1).

$$F_{in} = (\frac{\mu}{Y_{xs}} + m_s).V_r.\frac{c_{x0}}{c_s}.e^{\mu t} (4.1)$$

$Y_{xs}$ was overestimated to compensate for the effect of YE acting as a carbon source (Reiling et al., 1985). Exponential feeding was switched to a predictive feeding strategy based on offline glucose monitoring at 24 h. The carbon feed rate was manipulated to avoid glucose accumulation during this process. To avoid foaming during the process, 5 ml bolus of antifoam was added at regular intervals. The process was run for 71.5 h. The biomass profile exhibited a similar trend to that of Series A. The highest biomass achieved was 29.48 g/L at 31 h (Figure 4.4-a). It should be noted that due to absence of sampling event between inoculation and 31 h, biomass and EGN production trend can deviate from linearity. This is indicated by dotted connections between the two data points. Density of the dots increase after that, indicating closer approximation to linearity. Similar approximation can be applied to all the plots for this study. As expected, accumulated biomass continued to decrease over time. The highest EGN titer during cultivation was 86.06 %AU/L at 55 h (Figure 4.4-b). Cells stopped producing after that point and the titer was reduced to 76.17 %AU/L due to a dilution effect. Thus, scaling up with a factor of 1:150 reduced the EGN titer by 23%. An examination of specific productivity showed that the production rate increased during the cultivation until 27.5 h. After this point, it exhibited negative slope and was reduced to 12.38 %mAU/g/h (Figure 4.5). The biomass-specific growth rate at 12 h was 0.21 $h^{-1}$, followed by a gradual decrease to 0.004 $h^{-1}$ at 39.5 h. Cells eventually broke into the stationary phase, which was highlighted by a negative $\mu$ value after 39.5 h (Fig. 4.5). Better control was achieved for glucose profiles that did not accumulate beyond 3.94 g/L. However, acetate and lactate accumulation reached 9.83 and 13.2 g/L, respectively (Figure 4.6). Moreover, the accumulation of byproducts accelerated after 31 h. Notably, $q_{glucose}$ plotted over the time course for both Series A and B revealed that it began increasing when cells entered the death phase. Coincidentally, phosphate concentration was zero at that point despite its concentration in the batch media at

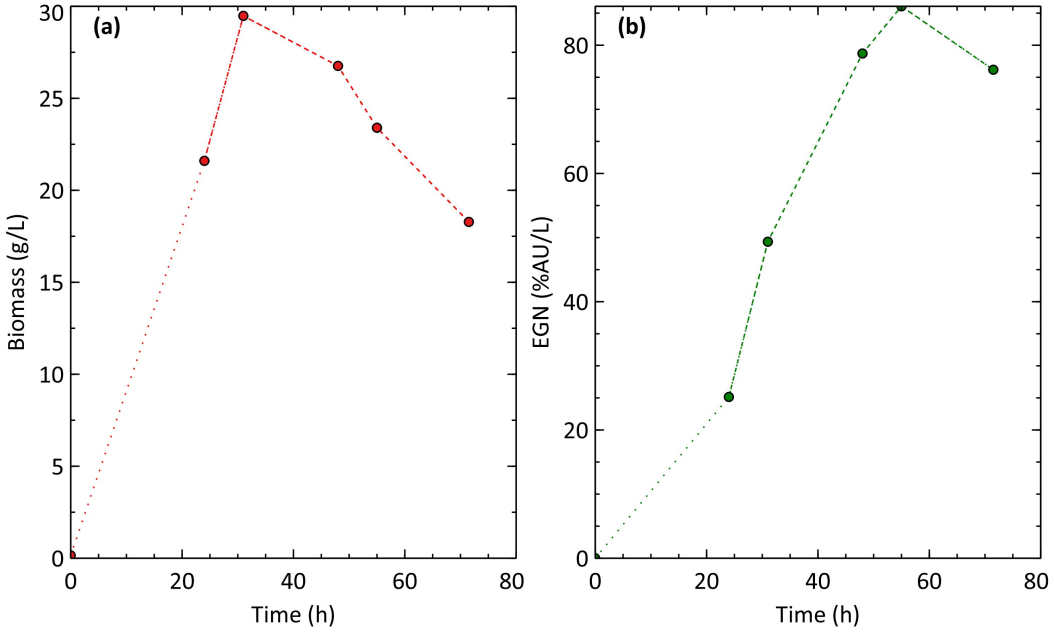



Figure 4.4: Cell density and the production of EGN for Series B. a: Biomass concentration in g/L; b: EGN titer in %AU/L during the time course of cultivation.

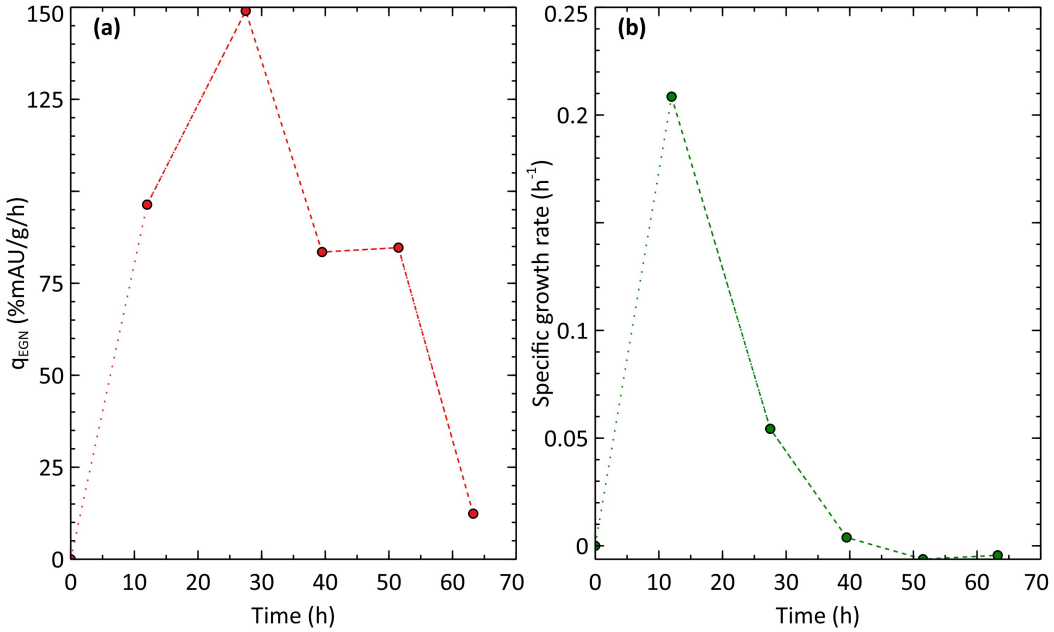



Figure 4.5: Production and growth performance for Series B. a: Biomass-specific productivity, $q_{EGN}$ in %mAU/g/h; b: Specific growth rate for EGN1, $\mu$ in $h^{-1}$.

the start of cultivation being 12.7 g/L. It was mainly supplied by potassium salts, while a small portion was supplied by YE (Christ & Blank, 2019). $q_{phosphate}$ was calculated, and it can be seen that as the productivity and specific growth rate decreased, phosphate consumption also decreased. This result establishes a direct correlation between product, biomass, and phosphate. Phosphate depletion also coincided with the local minima of $q_{glucose}$, after which glucose consumption dynamics were accelerated. The accumulation of lactate began around the same time as when phosphate was approaching exhaustion. Notably, phosphate is critical to the MEP pathway (Banerjee & Sharkey, 2014). Banerjee & Sharkey (2014) suggested that maximal MEcPP concentrations in chloroplasts are limited by the amount of available phosphate. *E. coli* biomass typically consists of 2–3% (w/w) phosphorus (Bauer & Ziv, 1976). Based on the biomass formed in Series B, we should not require more than 2.69 g/L of phosphate for growth. However, for an active MEP pathway, MEcPP is known to accumulate in *E. coli* and effuses to the extracellular environment if the pathway is unbalanced (Zhou et al., 2012). This might account for increasing phosphate demand for the EGN1 strain. We will use Series B study as a reference for MEcPP trend and MEcPP accumulation will be specified in percent, with 100% referring to Series B. Thus for Series B, close to 10 g/L of phosphate is drained as waste, which theoretically accounts for 100% of MEcPP and is equivalent to 268.04 %AU/L of EGN (assuming no accumulation of intermediates). Thus, the potential of the MEP pathway largely remains unexplored, with up to 300% flux diverted towards MEcPP production.

### 4.1.3 Salvation via phosphate feeding

To prove the critical importance of phosphate to the MEP pathway, we installed Series C—a study employing phosphate feeding in combination with glucose. Batch media composition was adapted from Series A and the study was performed in 2L Dasgip bioreactors. Phosphate was fed in addition to batch media provision. The accumulation of polyphosphates in *E. coli* can lead to ATP limitation caused by polyphosphate kinase (PPK) activity targeting the removal of excessive phosphate from the extra-

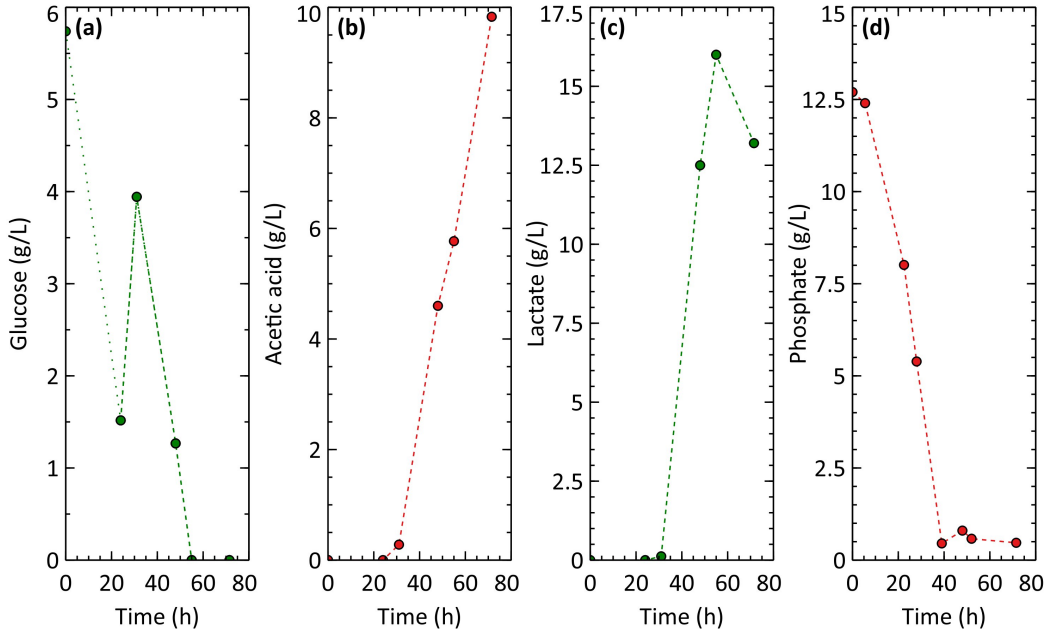



Figure 4.6: Substrate and byproduct formation for Series B. Starting from the left: a-Glucose concentration in g/L; b-Acetate concentration in g/L; c-Lactate concentration in g/L; d-Phosphate concentration in g/L.

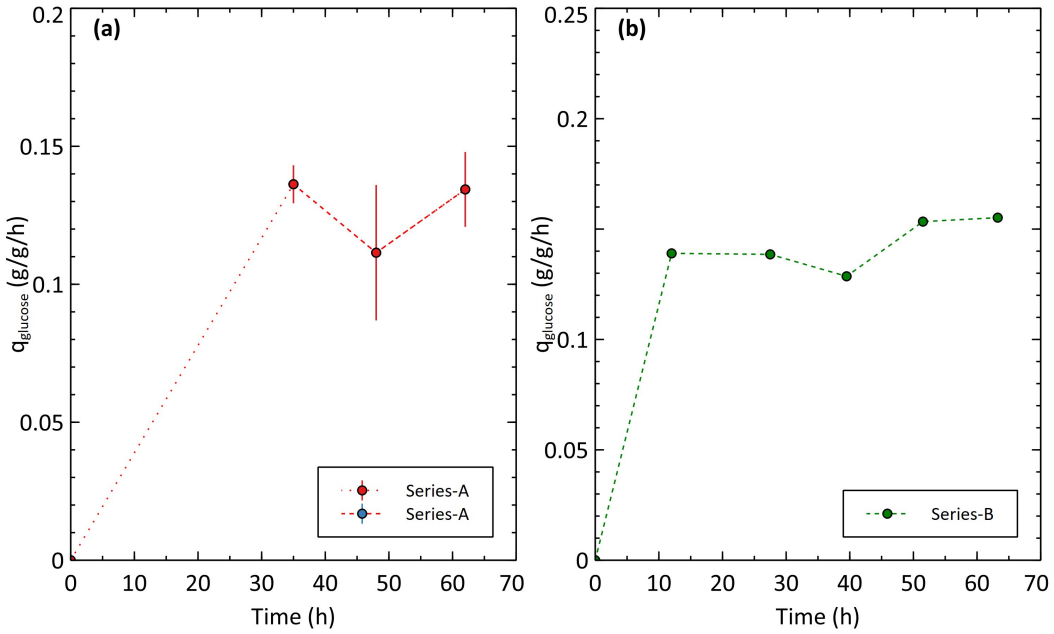



Figure 4.7: Substrate kinetics for Series A and B. Specific uptake rate of glucose for Series A (a, red) and Series B (b, green) in g/g/h.

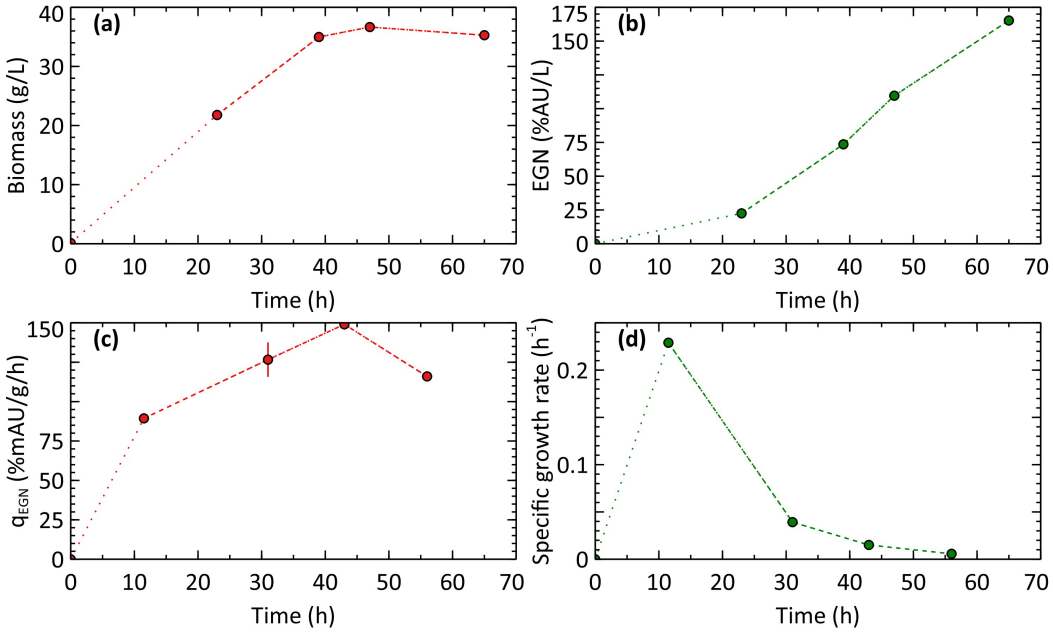



Figure 4.8: Cell density and the production of EGN for Series C. a: Biomass concentration in g/L; Top b: EGN titer in %AU/L; c: Biomass-specific productivity, $q_{EGN}$ in %mAU/g/h; d: Specific growth rate for EGN1, $\mu$ in $h^{-1}$ during the time course of cultivation.

cellular environment (Sharfstein & Keasling, 1994). Hence, increasing the phosphate concentration of the batch media was overlooked. The cultivation environment was preserved similarly to Series A. With the depletion of glucose in the batch media, feed-3 (see Section 3.2.6) including 35 g/L of phosphate was introduced. The highest biomass accumulated in the bioreactor was 36.66±0.66 g/L, which was 27% higher than in Series A (Figure 4.8). It should be noted that due to absence of sampling event between inoculation and 23 h, biomass and EGN production trend can deviate from linearity. This is indicated by dotted connections between the two data points. Density of the dots increase after that, indicating closer approximation to linearity. Similar approximation can be applied to all the plots for this study. The final biomass concentration was 35.28±0.66 g/L due to the dilution effect of the feed. The titer for EGN was increased to 165.22±3.72 %AU/L, representing a 65% increase when compared to Series A. Production was sustained throughout cultivation.

The specific productivity of EGN continued to increase during cultivation, peaking at 149.04±0.59 %mAU/g at 43 h, after which it decreased slightly to 115.99±0.92 %mAU/g at 56 h (Figure 4.8). The specific growth rate plot indicates that cells never

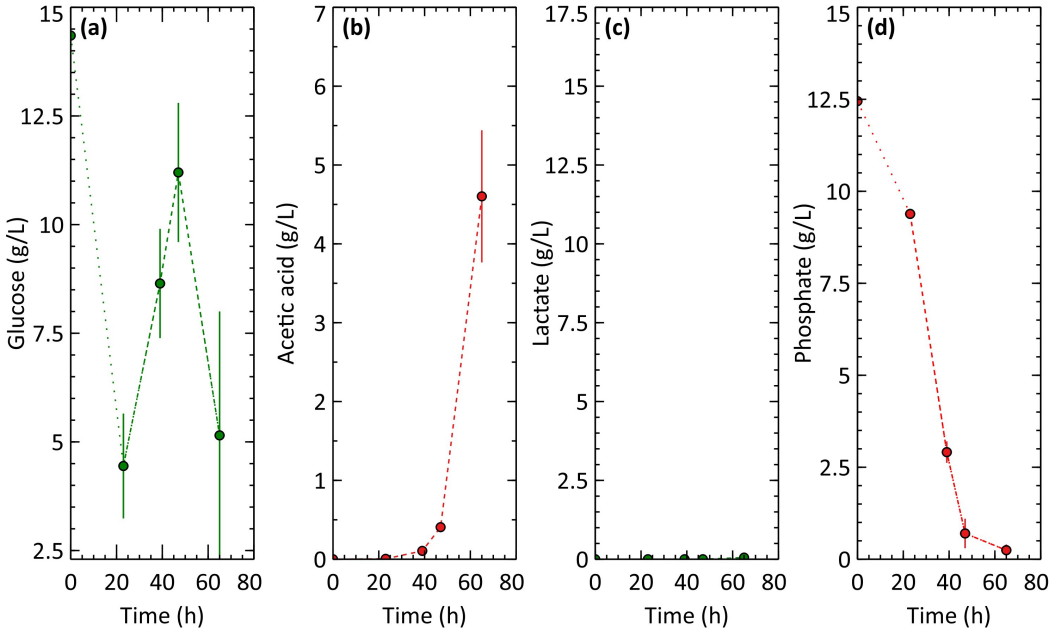



Figure 4.9: Substrate and byproduct formation for Series C. Starting from the left: a-Glucose concentration in g/L; b-Acetate concentration in g/L; c-Lactate concentration in g/L; d-Phosphate concentration in g/L.

entered the death phase since $\mu$ was always maintained above zero. This value was highest at 11.5 h with 0.23±0.00 $h^{-1}$ and then continued decreasing. However, this can be attributed to oxygen limitation in the bioreactor with agitation reaching the maximum value. Thus, the cells entered a microaerobic state at approximately 36 h. Although it was evident that the glucose feed rate control could not be tightened to avoid surplus glucose concentration in the bioreactor (Figure 4.9), the specific glucose uptake rate during cultivation did not increase in the latter part of the run, as seen in Series A (Figure 4.12). ATP synthase is inhibited when *E. coli* is subjected to phosphate limitation activity, which leads to a reduction in ATP energy charge. To cope with this, glucose uptake is increased (Schuhmacher et al., 2014). This explains the increased glucose uptake in the phosphate-limited/starvation phase in Series A and B. Together with oxygen limitation, glucose accumulation induced the accumulation of acetate, which reached as high as 4.60 ±0.83 g/L (Figure 4.9). However, no significant lactate was formed during the process. Lactate formation in *E. coli* can occur from pyruvate via the action of lactate dehydrogenase and the methylglyoxal pathway. A gene expressing lactate dehydrogenase, *ldhA*, was knocked out in EGN1. Hence, the

only plausible alternative could be the methylglyoxal pathway, which is active under phosphate-limiting conditions (Chakraborty et al., 2014). Under phosphate starvation, an increase in glucose uptake results in an excess of NADH formation due to upregulated metabolic activity. For Series A and B, phosphate starvation was induced in the oxygen-limited regime of the process. This, along with excess NADH, could inhibit the oxidation of lactate to pyruvate under the active methylglyoxal pathway. Therefore, lactate accumulation occurs.

With the addition of phosphate in glucose feed, the phosphate consumption rate at 43 h was 14.18±0.36 mg/g/h even though the phosphate concentration was less than 1 g/L in the bioreactor. With cells still in the growth phase, they were exposed to phosphate limitation, if not starvation. Hence, the feeding strategy requires further optimization. The excess phosphate added during Series C was 14.48 g/L based on the initial volume. With biomass amounting to approximately 5 g/L of phosphate, approximately 22 g/L of phosphate could have theoretically been drained to MEcPP. Based on Series B, 1 %AU/L of EGN accounted for 1.31% of MEcPP. Assuming linearity, 165.22 %AU/L EGN equates to 216.44% of MEcPP accumulation. $q_{phosphate}$ was highest when the cells were most productive. A decrease in biomass-specific uptake rate of phosphate explains the drop in productivity due to phosphate limitation.

### 4.1.4 Series D

For the thesis, we wanted to assess the criteria for scale-up via parallel tests. After addressing the issue of phosphate starvation, we scaled up the Series C process to a 300L bioreactor. Process conditions with reference to batch media composition, agitation cascade, aeration, and feeding strategy were identical to those used for Series B. Feed-4 (see Section 3.2.6), consisting of 600 g/L glucose and 45 g/L of phosphate, was installed after the exhaustion of the initial glucose.

The biomass attained in this study ultimately reached 31.05 g/L (Figure 4.10). However, after 55 h, it exhibited a significant drop that can be due to cell lysis. The bioreactor was in an oxygen-limited condition since the maximum OTR capacity of

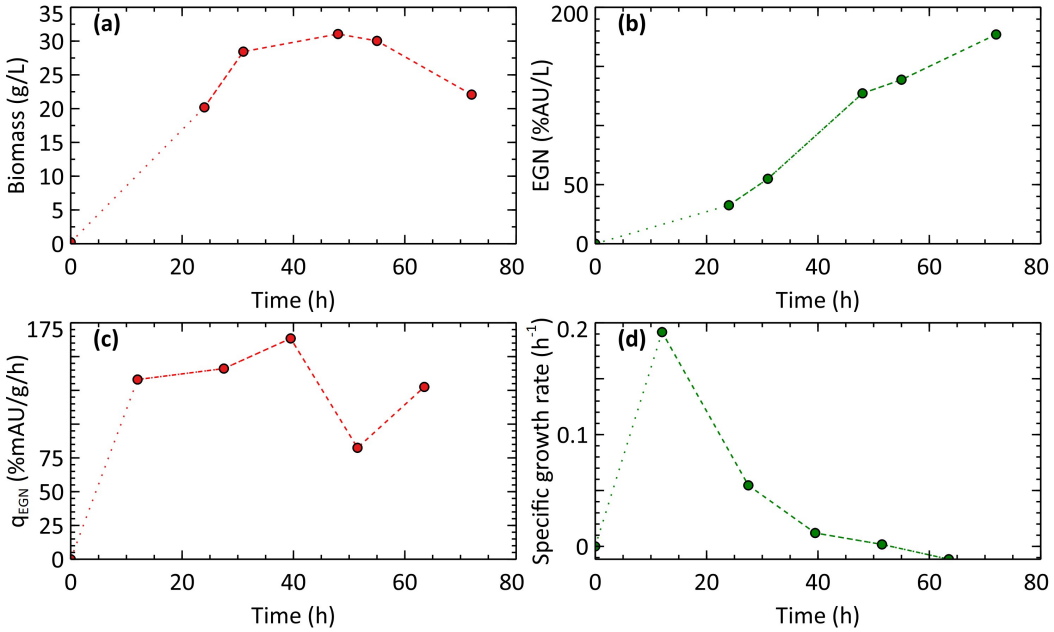



Figure 4.10: Cell density and the production of EGN for Series D. a: Biomass concentration in g/L; b: EGN titer in %AU/L; c: Biomass-specific productivity, $q_{EGN}$ in %mAU/g/h; d: Specific growth rate for EGN1 $\mu$ in h$^{-1}$ during the time course of cultivation.

around 95 mmol/min was already reached in the latter half of cultivation. Acetate was consequently formed, causing an inhibitory impact on biomass integrity (Soini et al., 2008). Although phosphate feeding resolved a part of the stability issue for cultivation, mass transfer limitation still resulted in a bottleneck at the large scale.

Cells were productive until the end of cultivation, reaching a maximum titer of 177.03 %AU/L. The specific productivity of EGN continued increasing until reaching 163.44 %mAU/g/h at 39.5 h. It should be noted that due to absence of sampling event between inoculation and 24 h, biomass and EGN production trend can deviate from linearity. This is indicated by dotted connections between the two data points. Density of the dots increase after that, indicating closer approximation to linearity. Similar approximation can be applied to all the plots for this study. Toward the end of cultivation, cells were producing at 127.53 %mAU/g/h. The specific growth rate profile was similar to that of Series C, with the highest value at 12 h (0.19 h$^{-1}$). Figure 4.11 indicates that total acetate accumulation at the end of the process was 8.99 g/L. As expected, no lactate accumulation was observed during the process, thereby consolidating the hypothesis on the effect of phosphate levels on lactate concentration.

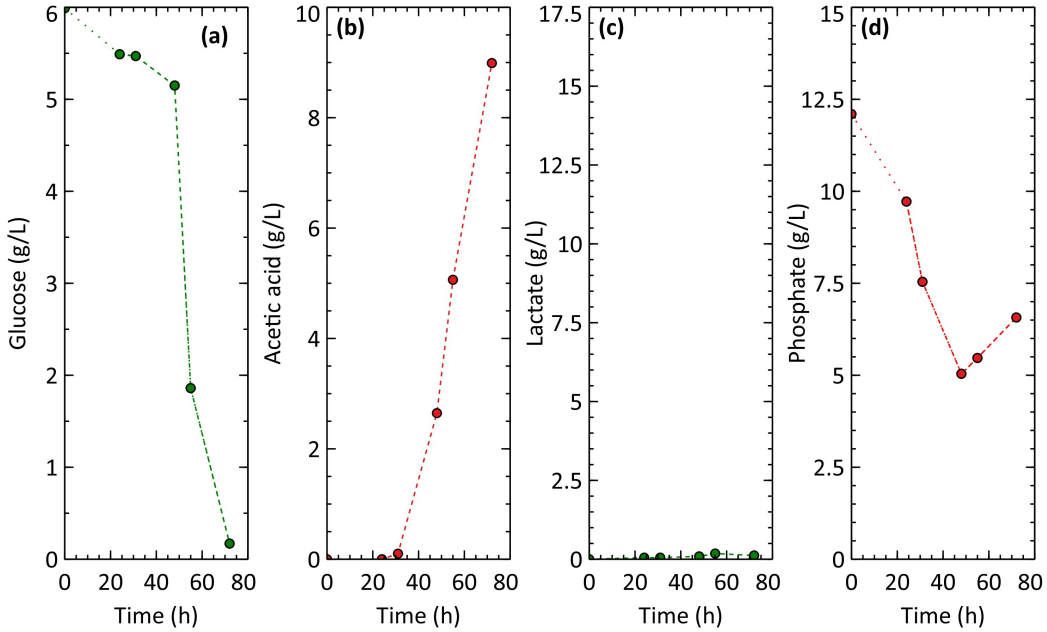



Figure 4.11: Substrate and byproduct formation for Series D. Starting from the left: a-Glucose concentration in g/L; b-Acetate concentration in g/L; c-Lactate concentration in g/L; d-Phosphate concentration in g/L.

While the process was successfully scaled up in terms of titer, biomass potential was still not reached in the cultivation, which showcases issues with the scale-up process. If oxygen limitation were not an issue, the EGN titer would have appreciated by 51.5% to 268.34 %AU/L assuming a biomass of 36 g/L being produced at 163.44 %mAU/g/h. The current process is far from lean, with around 7% glucose at disposal to acetate. Some outstanding issues include the development of an effective carbon limitation strategy and better scale-up criteria for the competent scalability of the process. Oxygen flux into microbial cells in aerobic cultivation strongly affects growth and product formation by influencing metabolic pathways and changing metabolic fluxes (Garcia-Ochoa & Gomez, 2009). For the MEP pathway, it also affects the availability of redox factors, which are important cofactors for most reactions involved in isoprenoid production (e.g., the NADPH-flavodoxin-flavodoxin reductase system) (Zhao et al., 2013). Iron-sulfur clusters are additional key enzymes for a productive MEP pathway. Hence, when aiming for better oxygen transfer, we must proceed with caution since excessive oxidative stress resulting from high DO levels can affect the iron-sulfur clusters (Lu et al., 2005). These concerns will be addressed in upcoming sections.

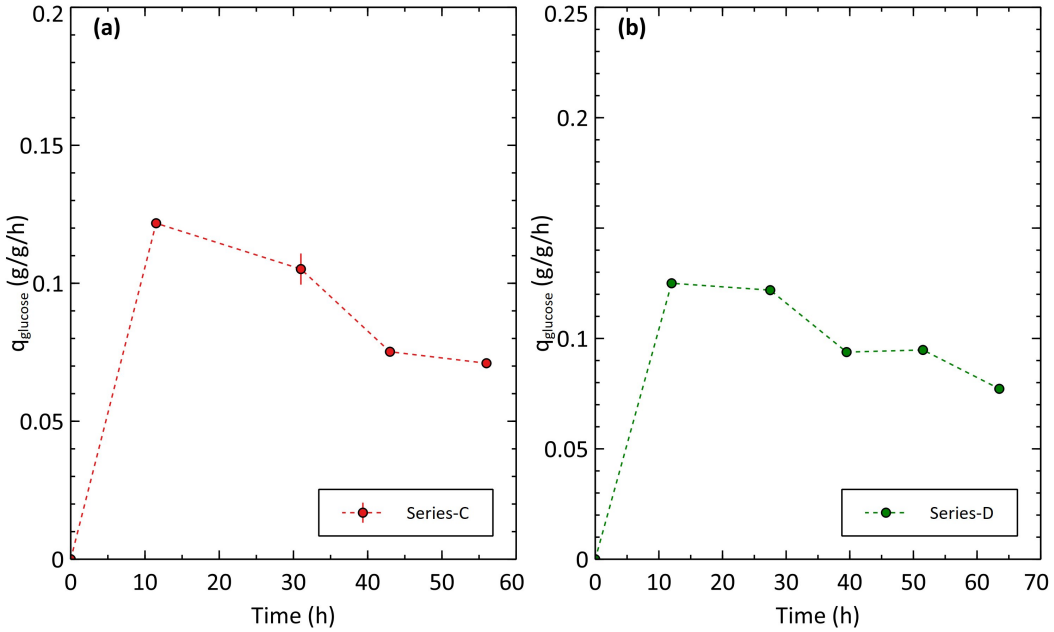



Figure 4.12: Substrate kinetics for Series C and D. Specific uptake rate of glucose for Series A (a, red) and Series B (b, green) in g/g/h.

## 4.2 The fundamental role of proper glucose and oxygen supply for producing terpenoids in recombinant *E. coli* showcasing amorpha-4,11-diene

One focus of this thesis was to understand the substrate kinetics of *E. coli* for strains upregulated with the MEP pathway genes. After determining the impact of phosphate on the EGN1 strain, we used that understanding to further explore the MEP pathway with the AMD strain. The AMD strain was engineered to produce amorpha-4,11-diene and was further modified to produce artemisinic acid. Efforts to decipher this interplay are outlined in (Patil et al., 2020).

While the MEP pathway is more carbon efficient than the MVA pathway, complexity in the regulations arises since the ATP cost for production could not be financed by excess NADH (Yang et al., 2016). The respiratory system of *E. coli* can be fine-tuned to generate ATP via well-balanced metabolism (J. Zhu et al., 2011). We focused on developing a fermentation process in *E. coli* by calibrating this balance.

Artemisinic acid and its precursor precipitate in the aqueous fermentation media at

high concentrations, thereby affecting mass transfer conditions within the fermenter and downstream separation (Westfall et al., 2012). To avoid this problem, an organic overlay of oil could be added to dissolve artemisinic acid. Since our ultimate focus was to develop a process for artemisinic acid, we supplemented the AMD4,11 production system with oil. The addition of oil affects the surface tension of the system and thus modifies air bubble breakage and the mass transfer coefficient (Dumont & Delmas, 2003). Featuring a direct correlation with respiration, we studied the impact of oil on AMD4,11 kinetics. We ran a 4-fold fed-batch fermentation (Series E). In Patil et al. (2020), our Series E study was described as "Series 1". Two different systems, namely monophasic and biphasic, were installed as duplicates in the experiment. With the exhaustion of glucose, carbon-limited feeding was installed with feed-2 (see Section 3.2.6) to subject all units to the same glucose feed rate in order to sequester the impact of the organic phase. The aqueous volume of the biphasic system reduced by 10%, thereby increasing the cellular glucose feed proportionally.

Notably, the monophasic system could accumulate more biomass than the biphasic system. The highest biomass achieved for the biphasic system (18.07±0.69 g/L) was 30% lower than that of the monophasic system (25.84±0.67 g/L) (see Figure 1a in Patil et al., 2020). While the AMD4,11 produced by the biphasic system at the end of cultivation was 0.56±0.02 g/L, this value was 6-fold lower than that of the monophasic system (3.46±0.11 g/L) (see Figure 1b in Patil et al., 2020). Series E reveals that no particular phenotypical changes for growth and AMD4,11 production were observed for the initial batch phase when nutrient consumption was apparently dominated by the utilization of YE (Suárez et al., 1998). The growth rate then decreased to 0.3±0.01 and 0.34±0.00 $h^{-1}$, reflecting the increasing impact of glucose consumption by feeding (see Figure 2d in Patil et al., 2020). Differences in biomass-specific glucose kinetics were clear, with the oil system exhibiting 20% higher biomass-specific glucose uptake rates than the oil-free system. Notably, volumetric glucose feed rates installed for the mono- and biphasic systems were the same.

Rising specific glucose uptake of the biphasic system concurred with the increased

biomass-specific oxygen demands and reduced growth rates of the AMD strain (see Figure 2b in Patil et al., 2020). Coincidently, during the initial part of the fed-batch process, $q_{AMD}$ decreased from 5.54±0.31 mg/g/h to 2.86±0.06 mg/g/h, eventually attaining a negative value. Notably, the $q_{O2}/q_{glucose}$ ratio was also higher in the biphasic system (see Figure 5 in Patil et al., 2020). FBA results illustrate the consequences of higher $q_{O2}/q_{glucose}$ on carbon availability at the pyruvate node for the biphasic system, which is also critical for MEP flux (Yang et al., 2016). More carbon seeps into the TCA cycle, thereby reducing the availability of pyruvate for the MEP pathway, as can be seen for B-16.75. Carbon is subsequently oxidized to carbon dioxide, thereby decreasing the yield for biomass and AMD4,11 on glucose. As soon as the oil system shifts into a microaerobic state (B-24), flux through tricarboxylic acid (TCA) is lowered by 96%. At this point, pyruvate formate lyase (PFL) becomes a more active option for the reduction of pyruvate. Carbon flux through pyruvate is partly diverted toward acetate, resulting in a 12-fold accumulation (see Figure 1c in Patil et al., 2020) for the biphasic system. The findings support the well-known property of *E. coli* to produce acetate once cells surpass a threshold for specific rates of glucose consumption (Eiteman & Altman, 2006). The presence of acetate complemented with oxygen limitation adversely affects the specific growth rate of *E. coli* (Luli & Strohl, 1990). During aerobic cultivation, oxygen is the preferred electron acceptor for transferring electrons from redox cofactors (e.g., NADH), leading to a stoichiometrically balanced formation of $NAD^+$, ATP, and $H_2O$. Notably, unwanted radicals and toxic compounds are the byproducts of this marriage with oxygen (Imlay, 2013). Accordingly, increased oxygen uptake rates require a proper supply of NADH, which is predominantly produced via TCA. Facing high NADH demand, cells may even make use of native transhydrogenase activity by converting NADPH to NADH (Spaans et al., 2015). The summation of all reactions involved in the production and consumption of NADPH resulted in a net NADPH flux for the FBA results. Interestingly, the net NADPH formation of growing cells was 17.6% lower with oil (B-16.75) than without (M-16.75), as shown in Figure 3 from Patil et al., 2020. The net NADPH flux availability was

17.7% lower for cells in the oil system when compared to those in the oil-free approach, which suggests a direct correlation with high biomass-specific uptake rates. Furthermore, AMD4,11 production accounted for 32.2% of net NADPH flux in the biphasic versus 18.9% in the aqueous system, which highlights the key finding of crucial NADPH availability for AMD production being hampered by high oxygen uptake rates caused by excessive glucose feeding.

Further evidence is provided by intracellular intermediates of the MEP pathway. Targeted metabolomics showed the accumulation of 1-deoxy-D-xylulose (DOX), a derivative of unstable DOXP, in the biphasic system. The latter is the product of the first pathway reaction condensing GA3-P and pyruvate to DOXP catalyzed by DXS (Sprenger et al., 1997). NADPH limitation created a bottleneck in the first step of the MEP pathway, thereby resulting in a 10-fold reduction in AMD4,11 production in the biphasic system (see Figure 1b in Patil et al., 2020)

The reduction of NADPH supply is caused by elevated NADH transhydrogenase activity consuming 69.3% and 81.9% of NADPH in the biphasic system versus 51.2% and 62% in the oil-free approach (calculated from all reactions involving NADPH). The resulting NADH entered respiration, which produced ATP and led to the doubling of formation rates (ATP4sr) (see Figure 3 in Patil et al., 2020). Flux for the ATP synthesis reaction (ATPS4r) was highest during growth in the biphasic system (33.47 mmol/g/h) and was reduced over time. Accumulating ATP pools in the biphasic system reflect this phenomenon (see Figure 3c in Patil et al., 2020). ATP concentrations diminished toward 50 h, mirroring the decreasing $q_{O2}$ and inefficient ATP synthesis due to microaerobic conditions (Chapman et al., 1971). Flux for the ATP synthesis reaction (ATPS4r) was highest during growth in the biphasic system (33.47 mmol/g/h) and reduced during fermentation.

Another interesting observation from the quantification of intracellular intermediates was the biomass-specific profile for MEcPP. It is a double phosphorylated intermediate that serves as a proxy for downstream bottlenecks in the MEP pathway. In particular, the need for doubled phosphorylation may provide insights into cellular phosphorus

management. Figure 3b in Patil et al., 2020 shows deviating intracellular MEcPP profiles for the mono- and biphasic systems starting from 26 h. Whereas control values steadily rise to 151.54±32.24 $\mu$M/g, MEcPP pools of the oil-containing tests appear to level out to 41.06± 6.36 $\mu$ M/g.

Arranging MEP flux for all FBA conditions revealed an inverse correlation to the ratio of $q_{O2}/q_{glucose}$. A summary of basic mechanisms is depicted in Figure 9 from Patil et al. (2020). Active volume in the bioreactor is reduced by the presence of oil, which increases the substrate availability per aqueous phase. This was more prominent during the fed-batch phase when the same feed rates of glucose were established for both systems. The increased $q_{glucose}$ cascaded rising $q_{O2}$ and positively affected flux through pyruvate dehydrogenase (PDH). Higher $q_{O2}$ required more NADH, which discounted NADPH availability and increased ATP pools. NADPH limitation and reduced levels of pyruvate decreased flux through the MEP pathway, thus discounting $q_{AMD}$. Interestingly, the observed phenomenon is particular to the MEP pathway and is likely to differ for the MVA pathway, which produces NADH (Yadav et al., 2012). Yang et al. (2016) observed that synergy between the MEP and MVA pathways can increase flux through the MEP pathway in NADPH-limiting conditions when both are expressed together in *E. coli*. This finding is congruent with the mechanism proposed in Patil et al. (2020).

Anticipating the critical importance of setting up a proper $q_{glucose}$ to $q_{O2}$ ratio for optimized MEP pathway activity, we conducted the Series F (Series 2 in Patil et al., 2020) study in the form of 4-fold fed-batch cultivations by installing a 10% lower glucose feed rate with the biphasic system to account for a reduced active volume. Encouraging results were observed, with both the systems following the similar biomass profile. However, the biphasic system led in production, reaching 3.47±0.14 g/L of AMD4,11, while the monophasic system could only produce 2.97± 0.1 g/L.

Lowering the glucose feed proportionally with the reduction of aqueous volume by oil addition (10%) substantially prevented unwanted performance loss. Briefly, trends of $q_{glucose}$ and $q_{O2}$ were overlapping (Figure 6a and b in Patil et al., 2020) with no po-

tential differences in NADPH and NADH availability. This is also reflected by equally similar courses of DOX, , $q_{AMD}$, and biomass-specific ATP contents. ATP levels remained between 0.5 and 1.5 $\mu$M/ g. Notably, DOX and MEcPP were accumulated during cultivation. However, as depicted in Figure 7 in Patil et al., 2020, relatively low maximum DOX levels were measured (approximately 0.035 $\pm$0.002 $\mu$M/g). Moreover, MEcPP rose to 128.31 $\pm$0.59 $\mu$M/g in the biphasic system. Since the flux to the MEP pathway was not limited by the low NADPH supply, amorpha-4,11-diene productivity was maintained in the biphasic system throughout cultivation, increasing AMD4,11 titer by 6-fold compared to Series E. Notably, MEcPP followed the same trend as AMD4,11 in all experiments, giving rise to the hypothesis that IspG activity was not sufficiently amplified.

Translating the lessons learned from Figure 9 of Patil et al. (2020) into optimized process conditions resulted in a 6.7-fold increase in titer production, resulting in a 3-fold increase in cellular productivity. Notably, there was a 6.54-fold increase in glucose-to-AMD conversion yield, while acetate accumulation fell by 67%. Tsuruta et al. (2009) noted the non-native MVA pathway in *E. coli* achieving 1.94 mg/g/h of amorpha-4,11-diene in a process lasting 160 h. In comparison, our approach reached 2.44 mg/g/h after only 60 h. In 2012, Westfall et al. (2012) attained an amorpha-4,11-diene production of 5.5 g/L in 95 h with a 3.23 C-mol% yield on glucose in *S. cerevisiae*. Upon extrapolating the volumetric productivity of our approach to 95 h, we would have reached approximately 5.2 g/L of AMD. Notably, upon installing apt $q_{O2}/q_{glucose}$ conditions, the *E. coli* AMD4,11 producer can potentially achieve 5.08% C-mol yields, which outcompetes the results using the MVA pathway.

### 4.2.1 Transcriptomics

Samples harvested from the Series E study were submitted for transcriptomic analysis. Cell pellets were collected at three time points: 9, 26, and 70.75 h. The biphasic system was referred to as Con1 and the monophasic as Con2. Notably, 70.75 h samples

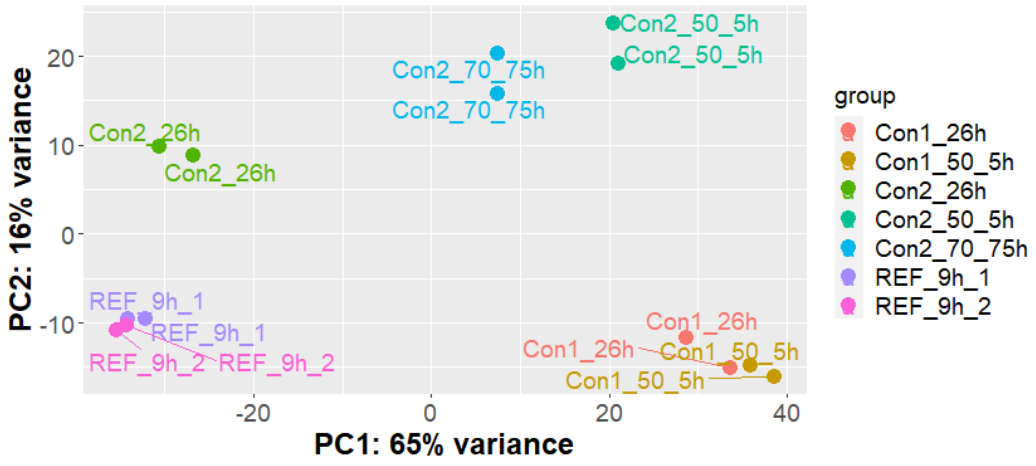



Figure 4.13: Principal component analysis shows clusters of transcriptome samples based on their similarity obtained at three process time points and two conditions. The variance of the data corresponds to the impact of principal components on the dataset.

were not available for the biphasic system since the sample integrity was compromised due to excessive foaming. This was further supported by the RIN numbers reported in Section 3.10.

Figure 4.13 shows the principal component analysis for the transcriptome samples. No difference was observed between two reference states (REF_9h_1 and REF_9h_2) belonging to 9 h samples. All cultivations had similar starting conditions. This result strongly supports the conclusions of Patil et al. (2020) by highlighting the impact of the fed-batch phase (starting after 11 h) on the system. Beyond 9 h, deviations could be seen between samples with oil (Con1) and without oil (Con2). Both transcriptional patterns can be divided into clearly separated clusters (up to 16 % variance).

Figure 4.14 shows the differentially expressed genes for both conditions. For the biphasic system (Con1), the gene expression of up- and downregulated genes exhibited an up to 10-fold increase when compared to the monophasic system as a part of long-term response. The increased up- and downregulation of genes in short-term response was

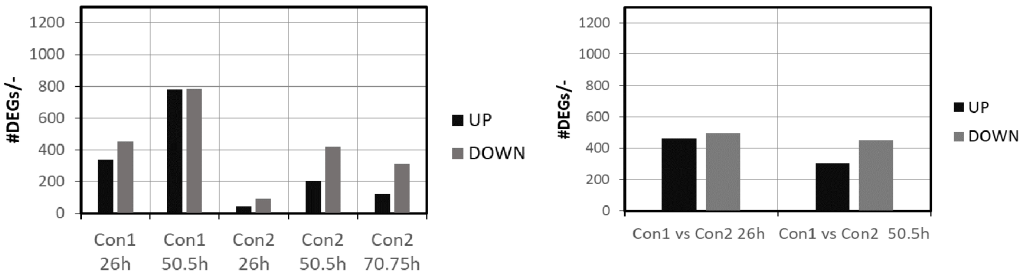



Figure 4.14: Number of DEGs whose expression was upregulated (black) or downregulated (gray). Left: Long-term response over all process times of Condition 1 (Con1) and Condition 2 (Con2) vs Reference (Ref 9h). Right: Short-term response of Con1 vs Con2 at corresponding time points.

also detected for the biphasic system. Thus, it can be concluded that the addition of oil affects transcriptome regulation.

To understand the biological response, different gene ontologies were analyzed for upregulation and downregulation. Biphasic systems seem to have a predominant upregulation of ribosome-related genes (Figure 4.16), while amino acid biosynthesis was downregulated for the biphasic system. Upon reviewing the response to substrate kinetics, it can be concluded that a higher availability of carbon substrate resulted in the downregulation of amino acid biosynthesis. After 26 h, many upregulated categories were downregulated after 50.5 h, and amino acid biosynthesis was still actively downregulated. Notably, the monophasic system showed the opposite regulation pattern after 26 h vs Ref_9h. It was observed that after 26 h, TCA was downregulated for the biphasic system, resulting in low NADH supply from the TCA cycle. Simultaneously, the biphasic system showed significant downregulation of MEP pathway genes, which was also evident in the experimental observation of Patil et al. (2020). The results from transcriptome analysis strongly support the observation made in the Series E study with respect to the biphasic system.

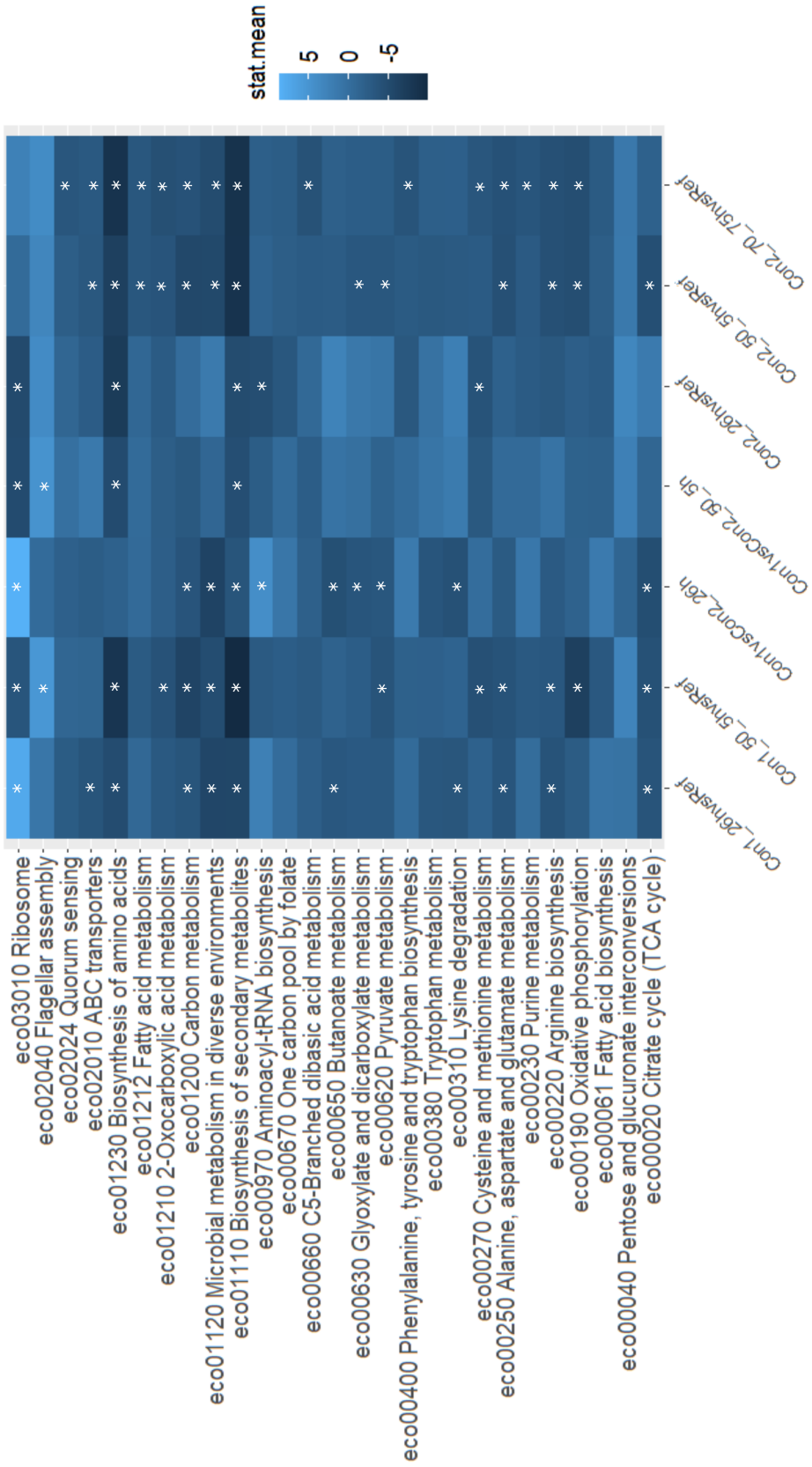



Figure 4.15: Top 27 KEGG Pathways with statistical means as the color scheme.

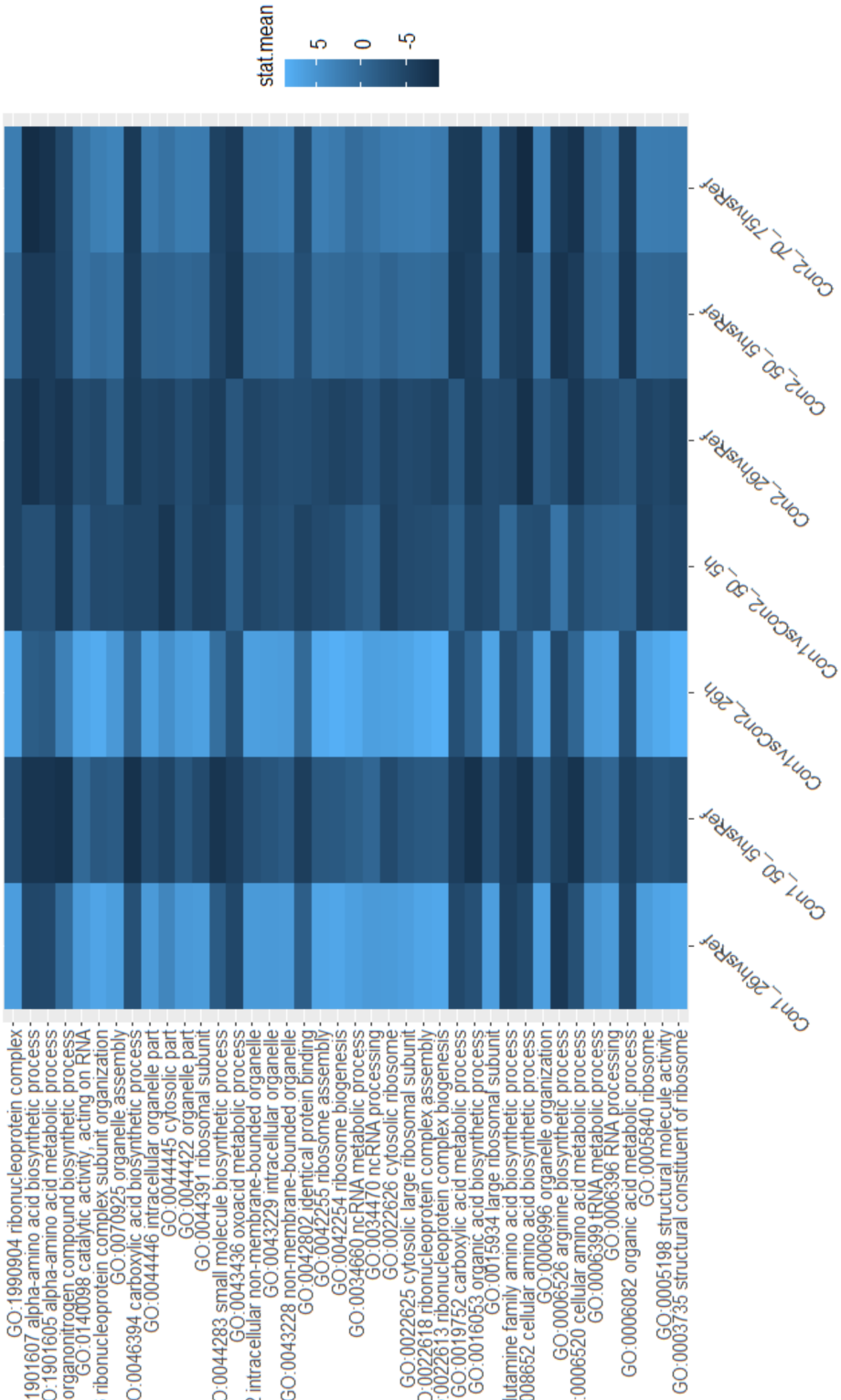



Figure 4.16: Top 40 GO categories over different sample comparisons with the statistical mean (stat. mean) as the color scheme.

It was also observed that envelope stress through oil (lipopolysaccharide biosynthesis and glycerophospholipid biosynthesis) are upregulated with oil. This might indicate enhanced/facilitated diffusion over the membrane. However, no clear evidence was seen regarding the perturbation of gene *cpxP*. Cells underwent the upregulation of flagellar assembly after 50.5 h in the biphasic system. This can be attributed to severe limitations encountered by the cell in the bioreactor environment. A closer look at terpenoid production via the MEP pathway indicated that gene regulation was not significantly altered for the biphasic system at 26 h. However, it was downregulated after 50.5 h (Figure 4.17)

We investigated transhydrogenase activity to understand the depletion of NADPH being correlated to DOX accumulation. It was evident that *pntA* and *pntB* were upregulated for 26 h in the biphasic system when compared to the reference. However, no such response was observed for the monophasic system (Figure 4.18). This coincides with the FBA result from Patil et al. (2020). Notably, *pntAB* is an active transhydrogenase during glucose metabolism (van Rijsewijk et al. (2016)). *arcA* also seemed to have been upregulated at 50.5 h, indicating the transition from aerobic to anaerobic conditions (Iuchi & Lin, 1993).

Transcriptomics also revealed that the regulation of gene expressions was related to oxidative phosphorylation. After 26 h for the biphasic system, the cytochrome *bd* complex was significantly upregulated. This might indicate high oxygen availability for ATP production (4.19). As indicated by the intracellular metabolomics data, transcriptomics results are similar to the ATP pools measured for the biphasic system. This explains the effect of increased specific oxygen uptake rate on terpenoid production.

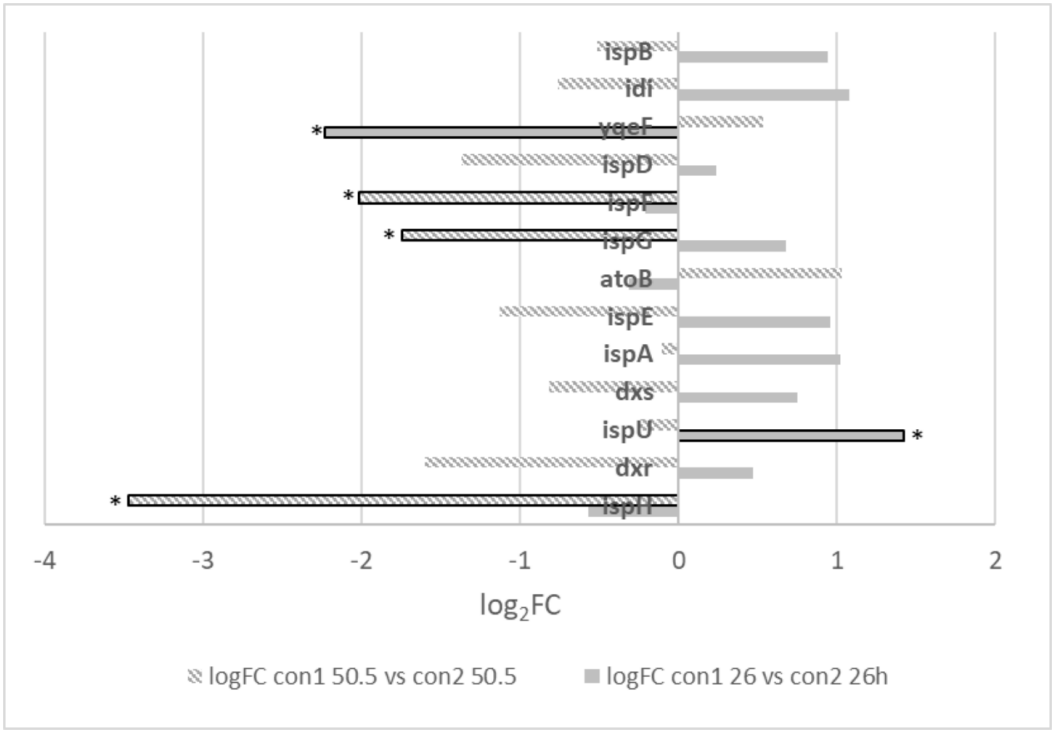



Figure 4.17: Regulation of terpenoid pathway genes at 26 h and 50.5 h for Con1 (biphasic) vs. Con2 (monophasic).

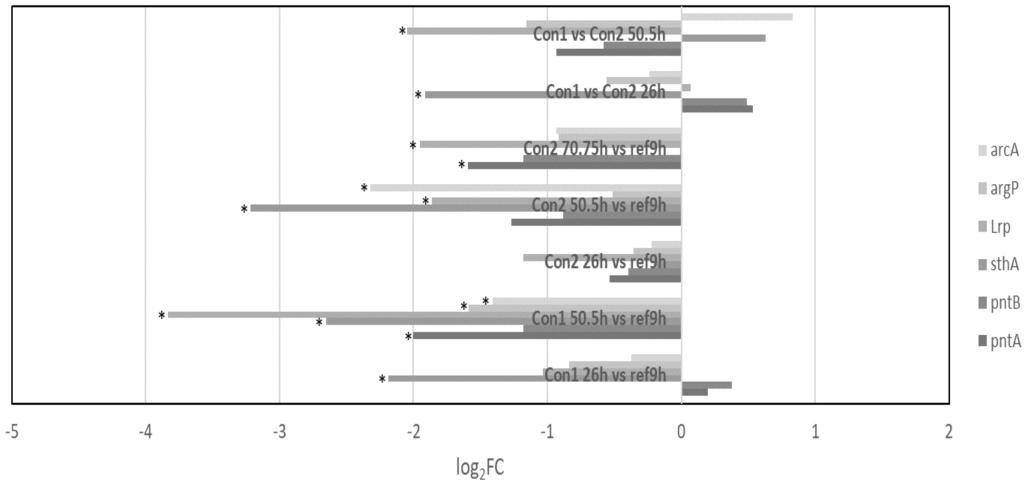



Figure 4.18: Transhydrogenase regulation based on the cutoff factor from van Rijsewijk et al., 2016 and adjusted for p-value<0.01.

| Genes | Con1_26h vs. Ref_9h | Con1_50.5h vs. Ref_9h | Con2_26h vs. Ref_9h | Con2_50.5h vs. Ref_9h | Con2_70.75h vs. Ref_9h | Con1 vs Con2 at 26h | Con1 vs Con2 at 50.5h |
|---|---|---|---|---|---|---|---|
| *cydA* | 1.37 | -1.12 | -0.1 | 3.27 | -0.61 | 1.38 | -4.60 |
| *cydB* | 1.40 | -1.23 | -0.16 | 3.15 | -0.50 | 1.40 | -4.54 |
| *atpA* | 0.07 | -2.66 | -1.23 | -1.67 | -1.73 | 0.92 | -1.37 |
| *atpB* | 0.37 | -2.80 | -1.05 | -0.97 | -1.17 | 1.11 | -2.13 |
| *atpC* | 0.13 | -2.37 | -0.91 | -1.60 | -1.60 | 0.69 | -1.13 |
| *atpD* | -0.19 | -2.98 | -1.18 | -1.86 | -1.76 | 0.63 | -1.48 |
| *atpE* | -0.09 | -3.15 | -1.29 | -1.51 | -1.71 | 0.71 | -2.14 |
| *atpF* | 0.44 | -2.49 | -1.22 | -1.27 | -1.57 | 1.28 | -1.61 |
| *atpG* | -0.32 | -3.03 | -1.22 | -1.75 | -1.69 | 0.62 | -1.56 |
| *atpH* | 0.18 | -2.01 | -1.20 | -1.14 | -1.32 | 1.02 | -1.24 |
| *nuoA* | -0.56 | -2.66 | -1.01 | -3.24 | -1.14 | 0.41 | 0.55 |
| *nuoN* | -2.81 | -2.67 | -1.14 | -3.10 | -1.95 | -1.93 | 0.17 |
| *sdhA* | -4.02 | -4.91 | 0.18 | -5.96 | -1.98 | -4.50 | 0.74 |
| *sdhB* | -4.19 | -5.57 | 0.01 | -5.87 | -2.37 | -4.40 | 0.10 |
| *sdhC* | -3.66 | -5.03 | 0.20 | -4.04 | -1.46 | -4.04 | -1.18 |
| *sdhD* | -2.82 | -4.10 | 0.11 | -4.51 | -1.18 | -3.24 | 0.10 |
| *cyoA* | -0.72 | -2.28 | -0.87 | -4.54 | -0.79 | 0.09 | 2.19 |
| *cyoB* | -1.15 | -3.03 | -0.76 | -4.58 | -0.82 | -0.56 | 1.37 |
| *cyoC* | -1.02 | -3.15 | -0.38 | -3.75 | -0.57 | -0.78 | 0.45 |
| *cyoD* | -0.61 | -2.99 | -0.16 | -3.70 | -0.53 | -0.50 | 0.66 |

Figure 4.19: Oxidative phosphorylation complex regulation based on cutoff values adapted from Erhardt et al. (2014) and adjusted for $p<0.01$. Significant values are colored.

For the biphasic system, we observed that transcriptome data laid out a good picture of gene expression for enzymes associated with the lower productivity of the MEP pathway.

### 4.2.2 Window of Optimum

Building upon the results of the previous section, we used the stoichiometric model to understand the behavior of AMD4,11 production over a wide range of specific oxygen and glucose uptake rates. A robustness analysis was performed on the model to predict the flux through the MEP pathway for AMD4,11 production corresponding to varying uptake rates of glucose and oxygen. We aimed to extend the findings of Patil et al. (2020). The aim of the analysis was to derive the optimum experimental conditions for achieving a specific productivity of more than 0.02 mmol/g/h of AMD4,11. The growth rate was set at 0.09 $h^{-1}$ and the results obtained are depicted in Figure 4.20. It can be observed that a global optimum exists for AMD4,11 production. Upon examining specific oxygen uptake rates for a given glucose uptake rate, it is evident that neither excessively high nor excessively low oxygen uptake rates are beneficial for productivity. There are regions in the response surface plot that have void solutions. These regions correspond to the solution for which ATP maintenance flux is less than 8.3 mmol/g/h. ATP maintenance reactions consider the consumption of ATP for non-growth-associated processes such as the maintenance of electrochemical gradients across the membrane (Orth et al., 2010). The highest possible productivity for AMD4,11 was calculated as 0.532 mmol/g/h, which is 26-fold greater than what was achieved by Patil et al. (2020). This highlights the unexplored potential of the MEP pathway. For a given oxygen uptake rate, AMD4,11 production was directly proportional to the specific glucose uptake rate. Thus, the glucose uptake rate could have been capped beyond 3 mmol/g/h. However, this would be far past the critical threshold rate for acetate formation (De Mey et al., 2007). Acetate formation can lead to severe stress and result in cell death (C. M. Takahashi et al., 1999). It was realized that the full potential of the system could be attained by subjecting cells to specific

glucose and oxygen uptake rates of 3 and 3.71 mmol/g/h, respectively.

To obtain a better understanding of the flux distribution at different nodes of Figure 4.20, we calculated the flux distribution of the model for five different constraints of glucose and oxygen uptake rate. These correspond to extreme or optimum solution constraints for the feasible solution space. The results of the simulation are presented in Table 4.1.

In Table 4.1, net NADH flux is calculated by summing up all of the reactions consisting of NADH as a substrate or product. It can be observed that PDH flux is minimal for an optimum solution, thus making more pyruvate available for the MEP pathway. If the process conditions tend to be at the oxygen limitation end ($q_{O2}$=2.0371 and $q_{glucose}$=3), PDH flux is highest when complemented with NADH excess, which would result in anaerobic fermentation products such as acetate. It is well known that if the capacity to oxidize redox factors is lower than glucose consumption, it can lead to NADH accumulation and result in acetate production (Han & Eiteman, 2019). On the other hand, if the cell kinetics correspond to a high specific oxygen uptake rate ($q_{O2}$=10 and $q_{glucose}$=3), ATP synthesis can be as high as 67.48 mmol/g/h. This would stand as a surrogate for high phosphorylation activity, resulting in the increased uptake of NADH. As a result, net NADH flux was lowest for this condition. Patil et al. (2020) observed that this would demand higher NAD transhydrogenase activity and the depletion of the NADPH pool. Since NADPH is an important cofactor for the MEP pathway, we would rightly observe lower AMD4,11 productivity for this condition. Thus, we could define an optimum window of approximately $q_{O2}$=3.71 and $q_{glucose}$=3 mmol/g/h (yellow region in Figure 4.20) for AMD4,11 production.

Notably, the MEP pathway intrinsically offers a well-equilibrated electron balance and the highest glucose-to-product yields to produce terpenoids. Accordingly, the aforementioned process settings may be qualified as key prerequisites to exploiting the full potential of the MEP pathway. We unraveled the complex metabolic interplay of glucose feeding, oxygen uptake, pyruvate, NADH, ATP, and NADPH supply. Optimum AMD4,11 production could be achieved by focusing on the key process parameters.

Table 4.1: Simulation results for window of optimum in mmol/g/h.

| $q_{O_2}$ | $q_{glucose}$ | $q_{AMD}$ | ATPSr | Net NADH flux | PDH flux |
|---|---|---|---|---|---|
| 10 | 3 | 0.209 | 67.48 | -2.55 | 2.74 |
| 3.7135 | 3 | 0.532 | 18.23 | -0.63 | 0.82 |
| 2.0371 | 3 | 0.02 | 10.8 | 6.68 | 4.49 |
| 3.7135 | 2.146 | 0.269 | 18.75 | -0.78 | 0.98 |
| 3.7135 | 1.387 | 0.0358 | 19.17 | -0.90 | 1.09 |

Although only shown for AMD, these findings are likely to be successfully applied for other MEP terpenoids since they address the fundamental problems of MEP-derived terpenoid production. In the sections to follow, we translate the findings from the AMD strain to the EGN2 strain.

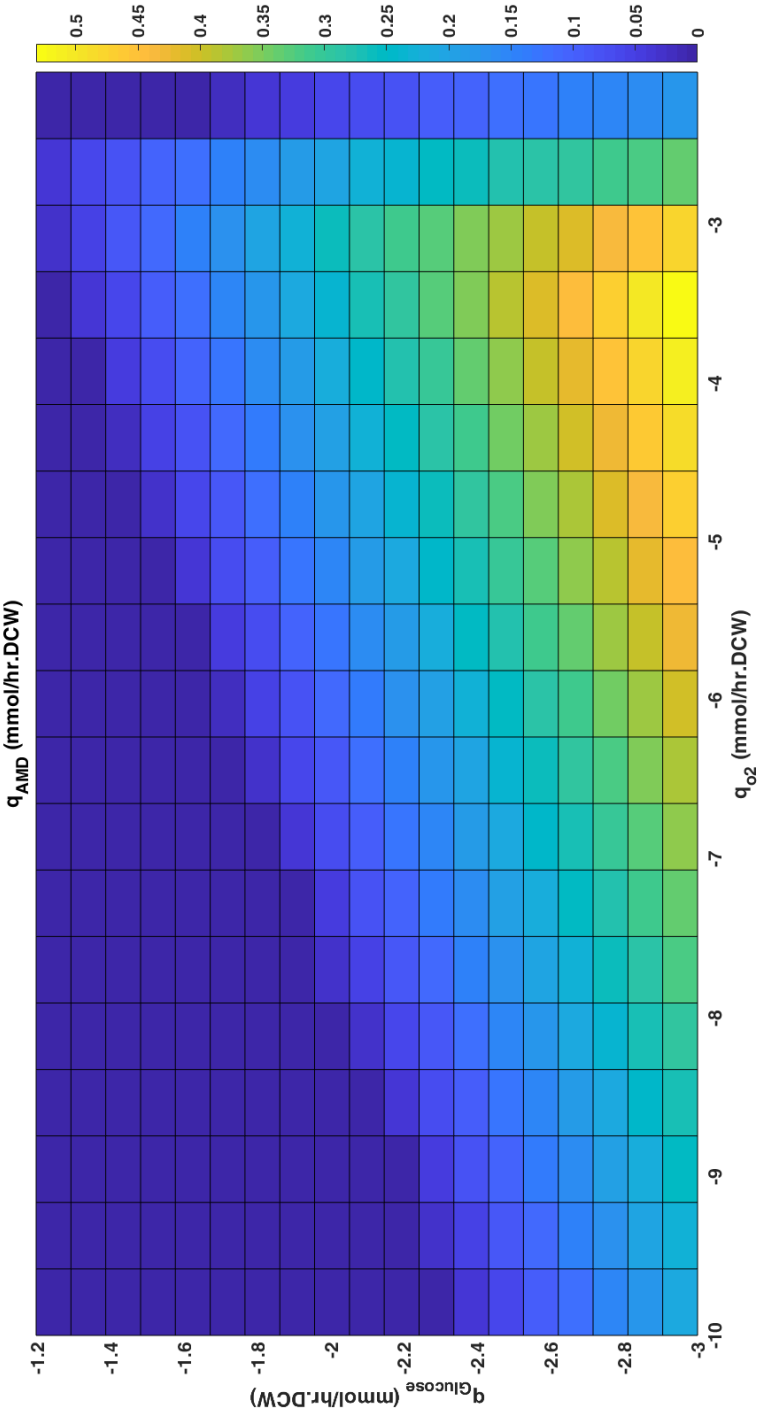



Figure 4.20: Results of robustness analysis for the developed stoichiometric model with $q_{AMD}$ designated as the objective function (values indicated in the colored bars) and $q_{glucose}$ and $q_{O_2}$ as the varying flux.

## 4.3 Effect of yeast extract on MEP flux-Series G

To further optimize the production of AMD4,11, we supplemented glucose feed with YE to lower the consumption of NADPH for biomass production in *E. coli* (Lyubetskaya et al., 2006). A positive impact was expected in this study. Process control for Series G is outlined as follows.

Cultivations in bioreactors were carried out in duplicate with the fed-batch mode and two different conditions were involved. Based on the findings from Section 4.2, control was established with the biphasic system establishing an appropriate glucose feed rate using feed-2 (see Section 3.2.6). Against this, we studied the implementation of feed-3 (containing 100 g/L of YE) in the fed-batch phase of the process. Process control was established as described in Patil et al. (2020). As soon as the base was switched from 6N $NH_4OH$ to 6N NaOH, an ammonium feed consisting of 99g/L of $(NH_4)_2SO_4$ and 18.75 ml/L of 25% $NH_4OH$ started to control the concentration of ammonia in the reactor between 30–60mM for the control condition. However, since YE was already acting as a source of nitrogen for the units fed with YE, ammonium feed was never pumped into them. Based on the glucose concentration in the bioreactor, the feed rate was adjusted to avoid the accumulation of residual glucose. This process ran for an estimated fermentation time of 69.5 h.

### 4.3.1 Yeast extract: Economizing NADPH

The highest biomass obtained during cultivation occurred when feeding YE to the bioreactors. As seen in Figure 4.21-a at 51 h, units with YE feeding could accumulate up to 35.00±0.38 g/L of biomass compared to the control (27.90±0.452 g/L). However, the biomass profile trends for both conditions were similar until the exhaustion of the initial glucose added to the batch media. Notably, both conditions were fed with the same amount of glucose. Thus, improvement in the yield of biomass on glucose for YE-fed reactors was observed with the value of 0.21 g/g compared to 0.13 g/g at the end of cultivation. Boosting the culture medium with complex media favors high cell density fermentation, recombinant protein yields, and acetic acid production

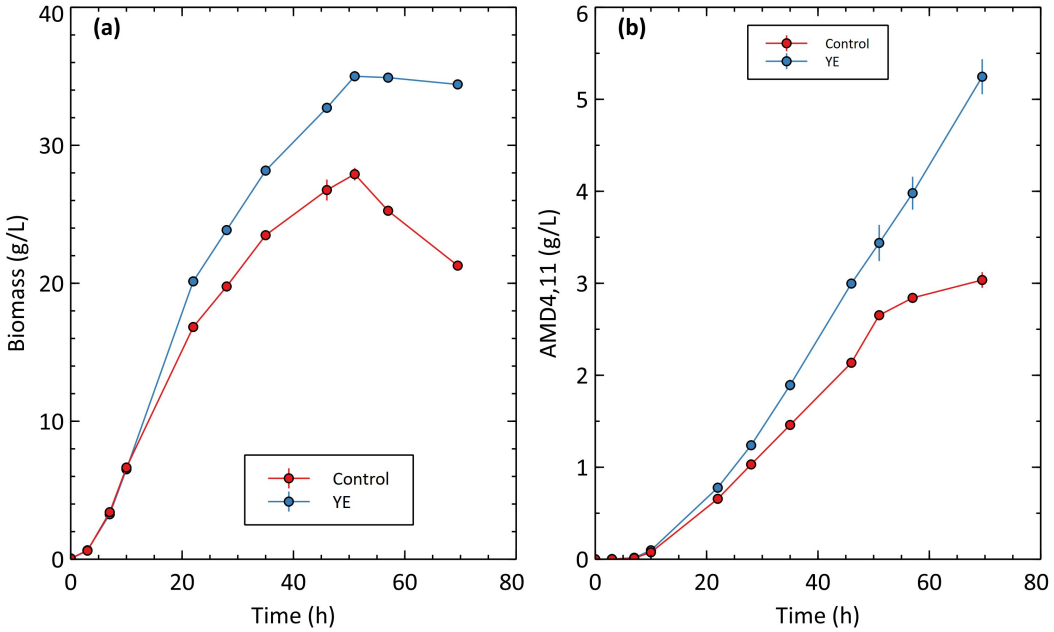



Figure 4.21: Cell density and production for Series G. Biomass concentration in g/l (a) and amorpha-4,11-diene titer in g/L (b) for control (red) and yeast extract (YE) feeding (blue).

(Riesenberg et al., 1991). Rothen et al. (1998) observed that addition of 0.4% YE saw an increase in the growth rate of *E. coli* by about 60% when compared to growth without YE in a continuous cultivation. For Series G, it was observed that cells never entered the stationary phase for YE feeding and continued growing until the end of the cultivation, albeit at a slower growth rate. Along with biomass, YE also complimented the production of AMD4,11, achieving a final titer of 5.24±0.19 g/L compared to 3.04±0.08 g/L for the control. Production entered a near-stationary phase for the control at approximately 51 h, while no such stagnation was experienced at any point in bioreactors fed with YE during cultivation. Specific growth rates were comparable during the batch phase for both conditions but began to deviate with the introduction of glucose feeding. $\mu$ was negative for the control after 54 h, indicating cell death and lysis. This can be attributed to acidic byproduct formation indicated by total base consumption for the process, which was 31% higher for the control. Biomass-specific productivity was sustained for the YE-fed group throughout fermentation, whereas it encountered a significant drop for the control after 48.5 h. Toward the end of fermentation, the control was producing at 1.45±0.13 mg/g/h,

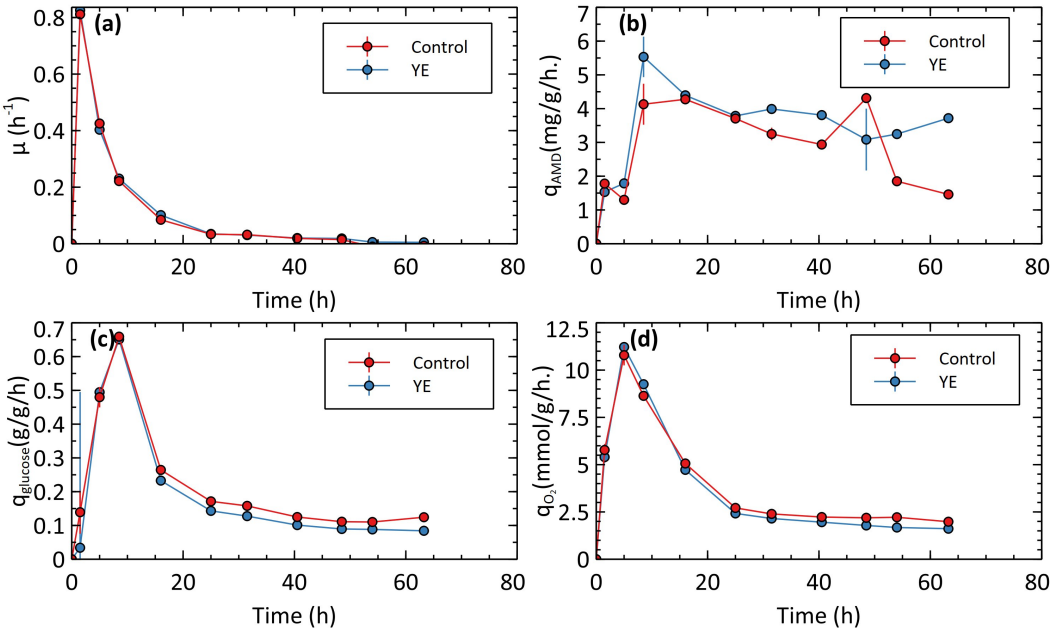



Figure 4.22: Biomass-specific dynamics for Series G. a: Biomass-specific uptake rate of glucose ($q_{glucose}$), b: oxygen ($q_{O2}$), c: productivity of amorpha-4,11-diene ($q_{AMD}$), and d: growth rate ($\mu$) for control (red) and YE-fed (blue) groups in Series G. Units: $q_{glucose}$, g/g/h; $q_{O2}$, mmol/g/h; $q_{AMD}$, mg/g/h.; $\mu$, $h^{-1}$.

while YE-fed bioreactors were producing at 3.72±0.09 mg/g/h.

While seeking to unravel the biomass-specific dynamics of glucose and oxygen, it was evident that throughout the fed-batch process, $q_{glucose}$ and $q_{O2}$ for YE-fed units were lower than the control. At 25 h, $q_{o2}$ was at 2.71±0.06 mmol/g/h for the control and 2.42±0.02 mmol/g/h for the YE-fed group. Deviation appears to be induced during the fed-batch phase. With lower $q_{O2}$, more pyruvate and NADPH were available for the MEP pathway in the YE fed group. Thus, they could be jointly responsible for the longevity of production. Notably, YE addition decreases the intensity of the citrate cycle. YE has a protein content of approximately 50%, of which approximately 20% is glutathione and 6% is nucleic acid. It is rich in 18 types of amino acids, functional peptides glutathione, dextran, mannan, trehalose, flavoring nucleotide, B vitamins, biotin, trace elements, volatile aromatic compounds, and other components (Proust et al., 2019). When amino acids are added to the medium, cells no longer need to synthesize certain substances, resulting in lower NADH production. Therefore, the intensity of the citrate cycle, which is one of the most effective sources of NADH, decreases (Lyubetskaya et al., 2006). Consequently, pyruvate pools are augmented for

utilization via other available reactions (J. Zhu et al., 2011).

Notably, YE addition is known to increase the production of acetic acid in *E. coli* (Suárez et al., 1998, Sigüenza et al., 1999). Organic acid formation can be signaled by the consumption of bases during cultivation. We experienced the reduced consumption of pH controlling base for our process, which points to the lower organic acid formation (acetate, in our case). One plausible explanation for this phenomenon could be the diversion of pyruvate toward MEP production in our strain. Hence, it is likely that this effect would not have been observed with the MVA pathway since the precursor requirement is focused on acetyl-CoA. Erian et al. (2018) reported that the BL21 strain of *E.coli* engineered for producing 2,4-butanediol (a product derived from pyruvate) produced 13.07 ±0.24 g/L 2,3-butanediol in the medium supplemented with YE as opposed to 0.44 ±0.10 g/L for the medium without YE.

However, the kinetics are more complex in the MEP pathway since it also involves the availability of G3-P for MEP flux. We will examine this in the following section.

## 4.4 Amino acid metabolism and relevant effects on the flux of the MEP pathway

To further understand the impact of YE on the growth and production of the AMD strain, we investigated the kinetics of quantifiable amino acids for the units fed with YE. Efforts have been focused on using a single or a few selected amino acids for the efficient and enhanced production of recombinant proteins in biotechnology (Y. Zhu et al., 1996). Based on the results from Series G, we conclude that YE had a serious impact on the biomass growth of AMD while simultaneously helping with the maintenance of AMD4,11 productivity (approximately 4 mg/g/h). Figures 4.23 and 4.24 shows the concentration profiles of all measurable amino acids during cultivation. We measured 16 amino acids using the method described in Section 3.6.4. The YE used in this study was abundant in L-alanine, L-glutamate, L-glycine, L-leucine, and L-tyrosine with concentrations ranging between 1500 and 2000 $\mu$M in the batch media.

Amino acids can directly synthesize biomass. However, if the catabolism of amino acid is active, we can predict it based on the buildup of intermediates involved in their degradation. This is complemented by the release of ammonia stemming from the activity of ammonia lyase transaminase (Eisenstein, 1991), which can increase the pH of the medium. During the batch phase, we observed that the pH of the bioreactor was increasing until 6 h, indicating the catabolic assimilation of amino acids along with their use for biomass buildup (Mikami et al., 2017). Concentrations for all amino acids either remained the same or decreased during the first 3 h of the process, except for glycine, isoleucine, lysine, and leucine, which indicates production. It has been observed that the cost of biosynthetic amino acids imposes a selective pressure to encode less costly amino acids in highly abundant proteins (Akashi & Gojobori, 2002). Thus, their consumption and composition in biomass should provide a good idea about the proteomes of the strain. Furthermore, the assimilation of low-cost amino acids leads to the production of some of the amino acids derived from them (i.e., glycine, isoleucine, lysine, and leucine in our case).

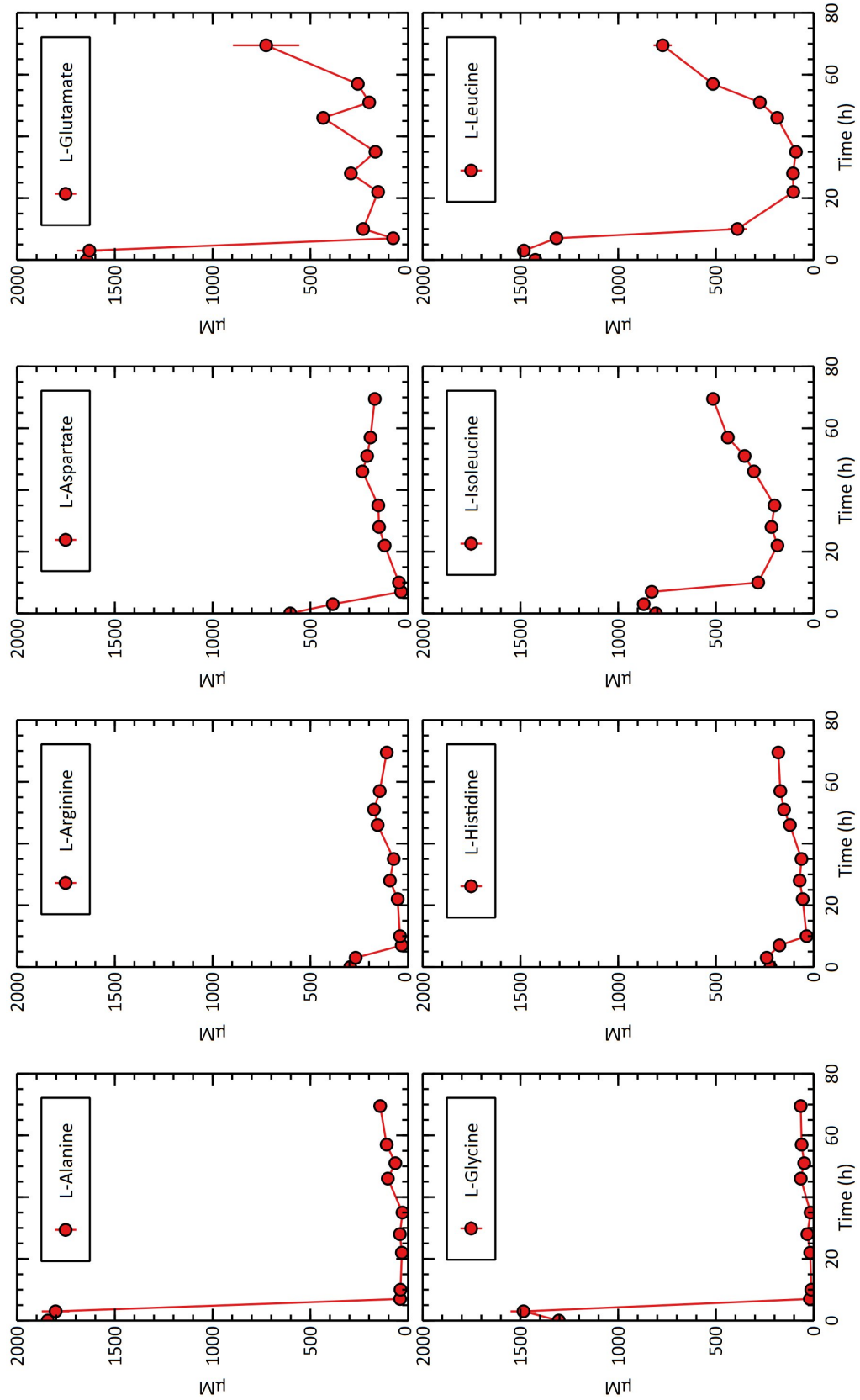



Figure 4.23: Concentration profile for amino acids in Series G-Part 1. Concentration of L-alanine, L-arginine, L-aspartate, L-glutamate, L-glycine, L-histidine, L-isoleucine, and L-leucine in $\mu$M for bioreactors with YE feeding.

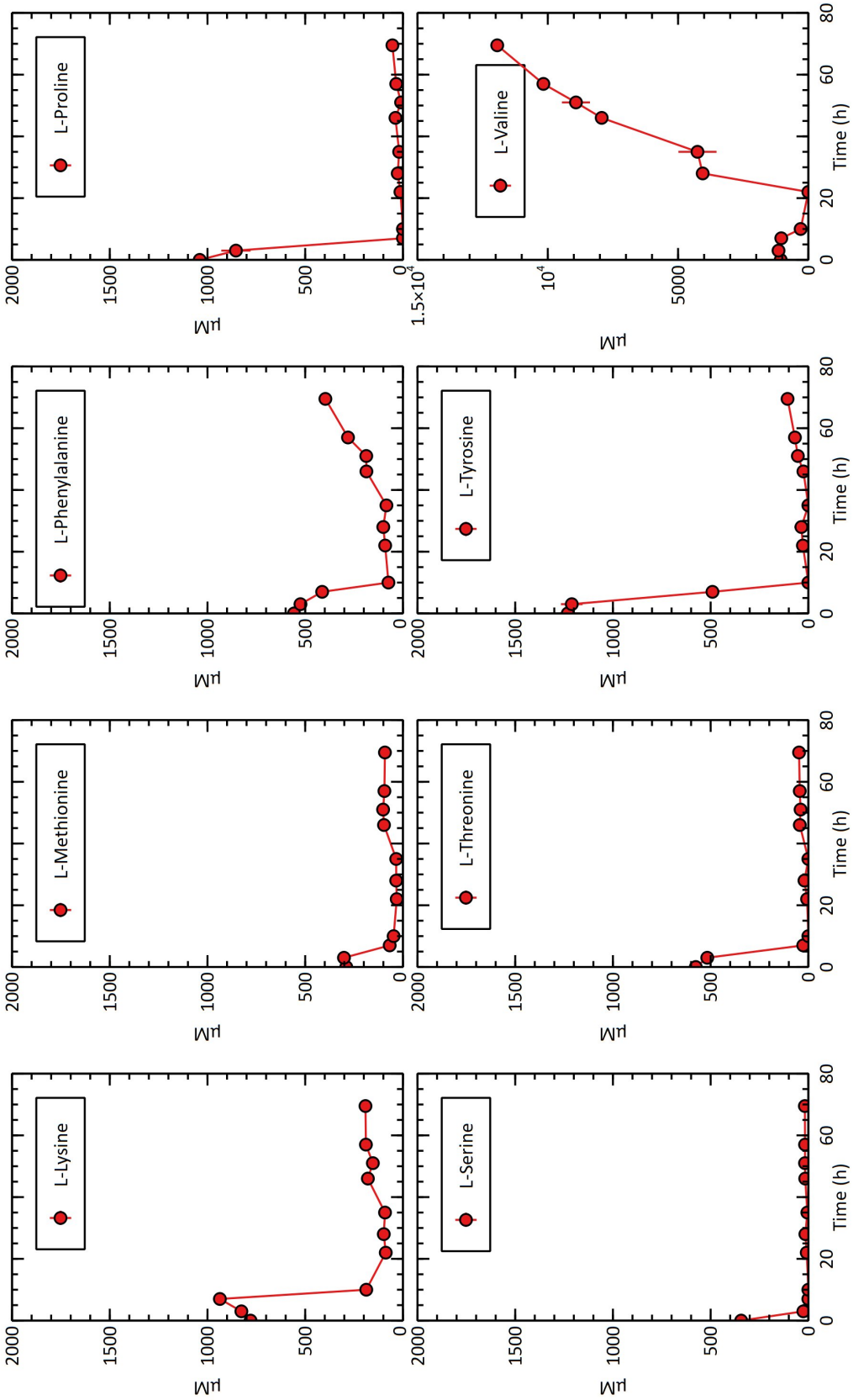



Figure 4.24: Concentration profile for amino acids in Series G-Part 2. Concentration of L-lysine, L-methionine, L-phenylalanine, L-proline, L-serine, L-threonine, L-tyrosine, and L-valine in $\mu$M for bioreactors with yeast extract feeding.

Figures 4.25 and 4.26 show biomass-specific amino acid uptake rates. Rates that fall on the higher side of the window include alanine, aspartate, glutamate, glycine, proline, and serine. Since these are low-cost amino acids, they should hypothetically be abundant in *E. coli* (Table 4.4). Energy cost is calculated based on glucose metabolism and is a measure of the number of high phosphate bonds involved in the production of a given amino acid (Akashi & Gojobori, 2002). During the batch phase, the highest growth rate of 0.825±0.01 $h^{-1}$ was observed at 1.5 h, where specific consumption rates of aspartate, proline, serine, arginine, threonine, and phenylalanine were operating at their highest $q_{aa}$ for the process. This explains the priority in consumption of these amino acids being targeted toward biomass production. However, the highest consumption rate was observed for L-serine (299.69±0.40 $\mu$mol/g/h) followed by L-alanine (226.59±10.47 $\mu$mol/g/h). This phase also coincided with a reduction in the specific glucose uptake rate, as seen in Figure 4.22. Then, aspartate is converted into oxaloacetate while serine and threonine are converted into pyruvate. Both these compounds seem to inhibit the phosphotransferase system for glucose uptake (Zampieri et al., 2019), which is precisely what we observed in our process. Once the insistent amino acids were consumed, glucose uptake was enhanced. Interestingly, the trend for L-valine was reversed during fermentation since it was consumed during the start of the fed-batch and then accumulated after 22 h. This can be due to the accumulation of pyruvate and assimilation of alanine, lysine, leucine, and isoleucine forcing dumping of the pyruvate pool in favor of valine synthesis.

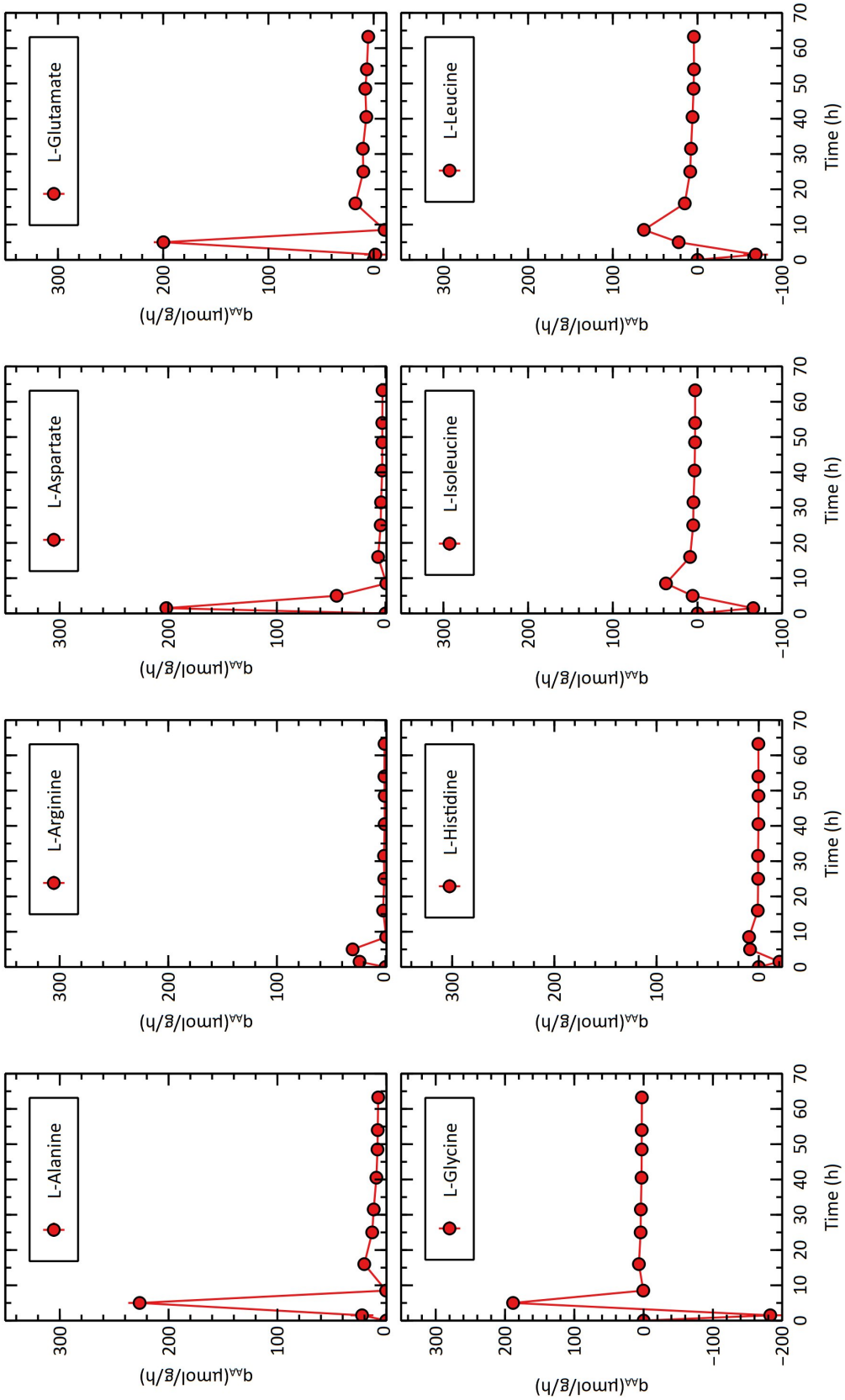



Figure 4.25: Biomass-specific kinetics for amino acid in Series G-Part 1. Biomass-specific uptake rates ($q_{AA}$) of L-alanine, L-arginine, L-aspartate, L-glutamate, L-glycine, L-histidine, L-isoleucine, and L-leucine in $\mu$mol/g/h for bioreactors with yeast extract feeding.

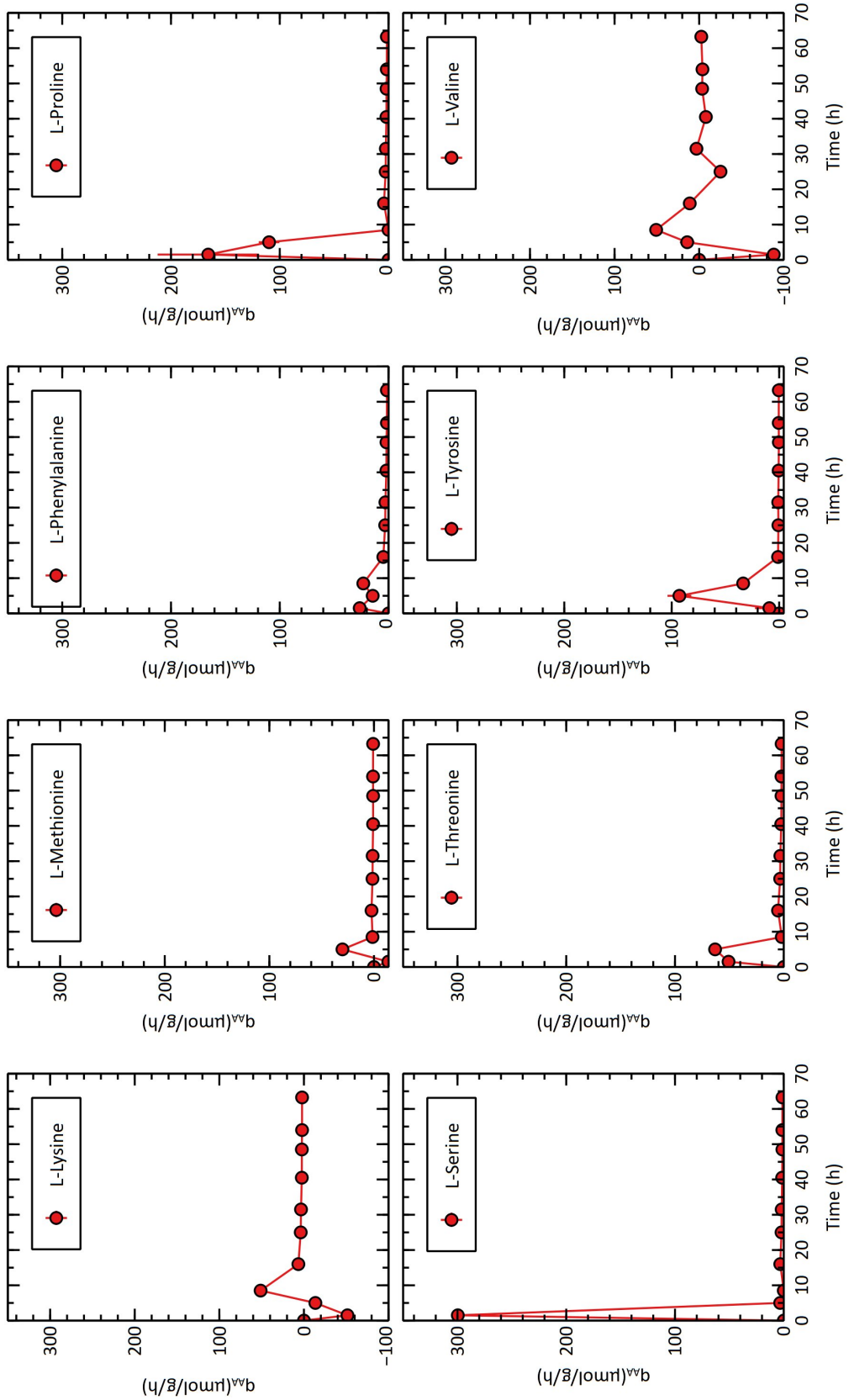



Figure 4.26: Biomass-specific kinetics for amino acid in Series G-Part 2. Biomass-specific uptake rates ($q_{AA}$) of L-lysine, L-methionine, L-phenylalanine, L-proline, L-serine, L-threonine, L-tyrosine, and L-valine in $\mu$mol/g/h for bioreactors with yeast extract feeding.

Table 4.4 shows the composition of *E. coli* biomass in terms of amino acids, as reported by Kaleta et al. (2013). We calculated the yield of amino acids in biomass, and it can be seen that $Y_{sx}$ had two maximas: one during the batch phase and another during the end of the fed-batch stage. We could divide the amino acids into two categories during the batch phase. As reported in Table 4.4 during the batch phase, alanine, aspartate, glutamate, proline, serine, and tyrosine were all above the $Y_{sx}$ value, indicating their catabolic degradation. Except for tyrosine, all of these belong to the lower end of the energy cost index. However, tyrosine peaked in consumption toward the latter half of the batch phase when glucose consumption had already peaked. The remainder were lower than the reported $Y_{sx}$ value, which highlights consumption to produce biomass directly and no degradation.

Upon moving into the fed-batch phase, the $Y_{sx}$ value for all of them either decreased or remained the same, which strongly indicates glucose metabolism to fulfill carbon demand. However, toward the end of cultivation, all amino acids except for arginine, glycine, histidine, serine, and tyrosine had higher values of $Y_{sx}$ than the threshold value. Notably, the $Y_{sx}$ values for serine and tyrosine were lower than their maximas in the batch phase but were close to the threshold values shown in the Table 4.4. The degradation of amino acids results in the increased accumulation of pyruvate toward the end of the process, which could result in sustained productivity. Simultaneously, this indicates glucose limitation in the process and salvation by amino acid catabolism. Since the ammonia concentration was also maintained above 1 g/L, the possibility of ammonia limitation can be discarded. The reasons for glucose limitation could either be the accumulation of oxaloacetate, pyruvate, or both. This also correlates with the accumulation of valine during the process. While more work needs to be directed for this analysis, such work extends beyond the scope of this thesis.

To further investigate the impact of amino acids on growth and productivity, we performed an FBA on the iJR904 model of *E. coli* (modified as described in Patil et al. (2020)) with the objective function of the specific growth and specific productivity

Table 4.2: Energy costs for the synthesis of amino acids. Metabolic energy costs for amino acids calculated in terms of the high-energy phosphate bonds involved in their production from the metabolism of glucose (Akashi & Gojobori, 2002).

| **Amino acid** | **Energetic cost** |
|---|---|
| Ala | 11.7 |
| Gly | 11.7 |
| Ser | 11.7 |
| Asp | 12.7 |
| Asn | 14.7 |
| Glu | 15.3 |
| Gln | 16.3 |
| Thr | 18.7 |
| Pro | 20.3 |
| Val | 23.3 |
| Cys | 24.7 |
| Leu | 27.3 |
| Arg | 27.3 |
| Lys | 30.3 |
| Ile | 32.3 |
| Met | 34.3 |
| His | 38.3 |
| Tyr | 50 |
| Phe | 52 |
| Trp | 74.3 |

Table 4.3: Amino acid consumption of biomass for *E. coli*. Number of moles of each amino acid per unit biomass ($Y_{sx}$) in $\mu$moles/g (Kaleta et al., 2013).

| **Amino acid** | $Y_{sx}$ |
|---|---|
| Ala | 514.78 |
| Arg | 293.37 |
| Asp | 243.55 |
| Glu | 265.69 |
| Gly | 608.88 |
| His | 94.10 |
| Ile | 287.83 |
| Leu | 448.36 |
| Lys | 343.19 |
| Met | 154.99 |
| Phe | 177.13 |
| Pro | 221.41 |
| Ser | 215.88 |
| Thr | 254.62 |
| Tyr | 138.38 |
| Val | 420.68 |

of AMD4,11 in two different simulations. Simulations were compiled with the specific glucose uptake rate constrained at -1.31 mmol/g/h, the lower limit of $\mu$ set at 0.0823 $h^{-1}$, and ATP synthesis rate constant at 8.3 mmol/g/h. The objective function was flux through reaction Ex_AMD converting FPP to AMD4,11. These constraints were determined from the Series E study based on a simulation for M-16.75. For the optimization of the specific growth rate, no production condition was implemented. It was observed that for the amino acid consumption rate of 0.8 C-mol/g/h, only aspartate, glutamate, serine, glycine, threonine, arginine, alanine, and proline has a positive impact on the specific growth rate. Notably, proline exhibited an outstanding specific growth rate of 0.1055 $h^{-1}$ when compared to 0.0925 $h^{-1}$ for the control condition metabolizing only glucose. Similar characteristics were observed for the specific productivity of amorpha-4,11-diene with the addition of individual amino acids to the FBA. However, while simulating optimized production, the minimum growth rate was constrained to 0.0823 $h^{-1}$. Again, proline outclassed all other amino acids with a specific productivity of 0.066 mmol/g/h. Negative impacts among the remaining amino acids can be explained by their position at the higher end of the energetic cost spectrum.

## 4.5 Process transferability

Once we established the interplay between substrate dynamics for AMD4,11 production, we wanted to implement the same principle to attempt the transfer of process for EGN production. This was achieved by installing a fed-batch process with the EGN2 strain. The strain was an upgrade to EGN1 and exhibited better performance during the strain screening process carried out by Manus Bio Inc. (Cambridge, USA). EGN2 had overexpression of the genes *ispG* and *ispH*, which resulted in the lower accumulation of MEcPP. Based on the process developed for AMD4,11, we addressed the issues of oxygen limitation that can lead to acetate accumulation. The further optimization of glucose uptake would affect the specific oxygen uptake rate by eventually optimizing

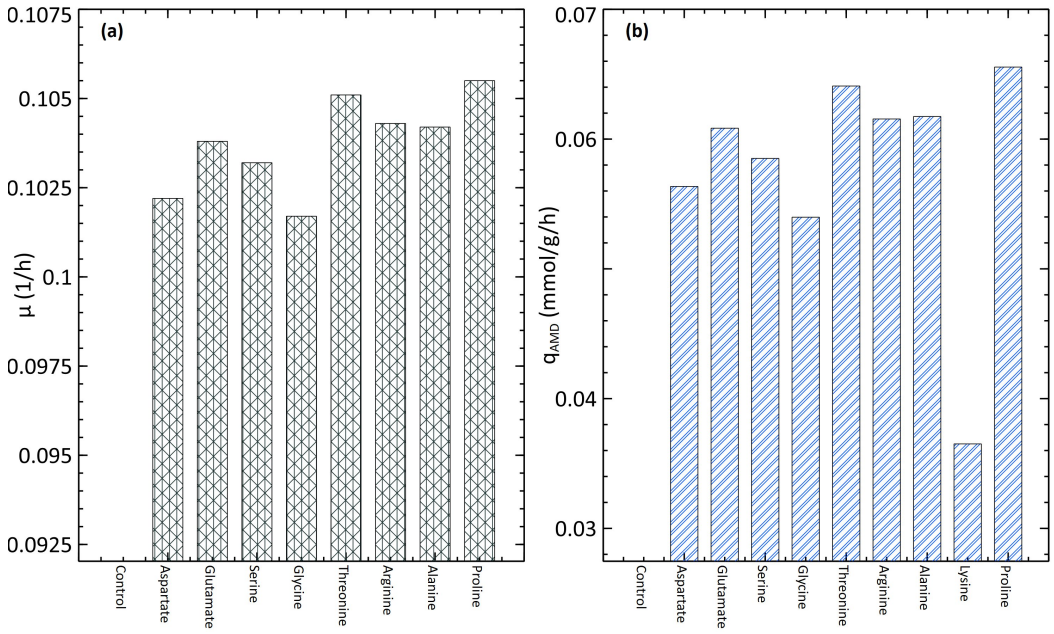



Figure 4.27: FBA results for amino acid consumption for the AMD strain. a: Simulation 1 with specific growth rate $\mu$ as an objective function; b: Simulation 2 with $q_{AMD}$ as an objective function and $\mu$ constrained at 0.0823 $h^{-1}$.

the flux through the MEP pathway. The following section describes the Series H study.

### 4.5.1 Process control

Process control for the fermentation was the same as for Series F. $p_{O2}$ was maintained at 40% of air saturation until 40 h, after which it was reduced over 6 h to 15%. An agitation cascade (300–1200 RPM) was used to control the $p_{O2}$ levels. This was also a biphasic system with 10% safflower oil added as an overlayer to capture EGN produced during fermentation. Process control was established as described in Patil et al. (2020). Cultivation was terminated after an estimated fermentation time of 72 h.

### 4.5.2 Successful handover

The highest biomass achieved during fermentation was 42.13±1.66 g/L at 72 h (Figure 4.28-a). Biomass growth was maintained at a positive rate throughout the cultivation. The EGN titers obtained were 292.55±1.59 %AU/L at the end of the cultivation

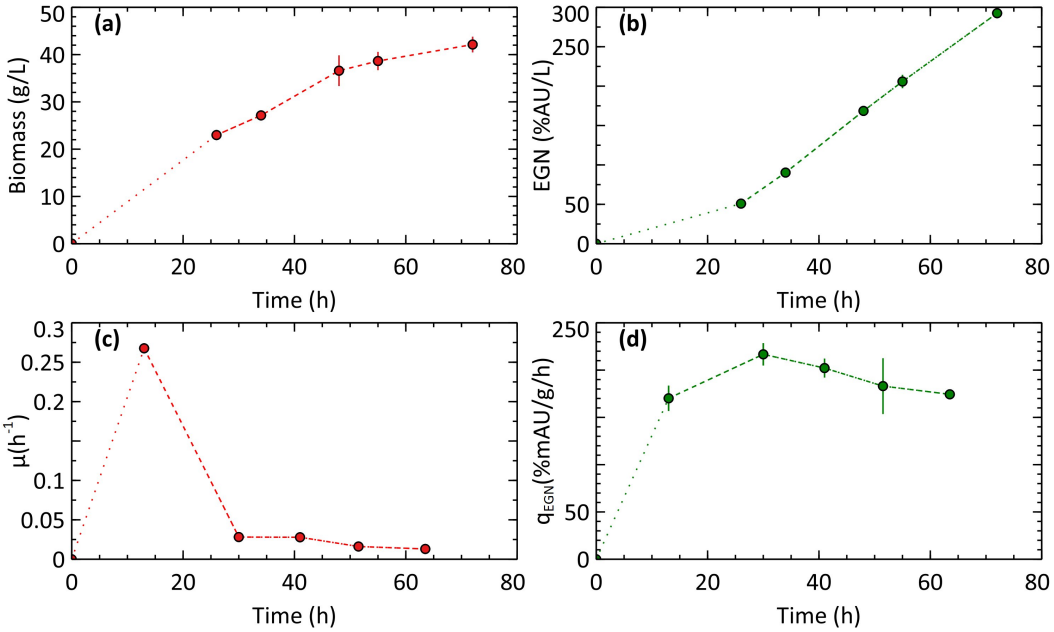



Figure 4.28: Cell density and the production of EGN for Series H. a: Biomass concentration in g/L; b: EGN titer in %AU/L; c: Biomass-specific productivity, $q_{EGN}$ in %mAU/g/h; d: Specific growth rate for EGN2, $\mu$ in $h^{-1}$ during the time course of cultivation.

(Figure 4.28-b). It should be noted that due to absence of sampling event between inoculation and 26 h, biomass and EGN production trend can deviate from linearity. This is indicated by dotted connections between the two data points. Density of the dots increase after that, indicating closer approximation to linearity. Similar approximation can be applied to all the plots for this study. Thus, the issue of process breakdown was addressed. EGN titer corresponded to 57.39 mM of IPP formation (theoretically). The best amorpha-4,11-diene titer obtained on glucose feeding was 16.9 mM, which corresponded to 50.7 mM of IPP formation.

Biomass-specific growth rate followed a similar trend to AMD. The highest value of 0.27±0.00 $h^{-1}$ was observed at 13 h, after which it began decreasing due to glucose and oxygen limitation. The specific productivity of EGN was maintained above 170 %mAU/g/h throughout the entire process (the highest productivity to date). The highest production rate was observed at 30 h (216.69±11.79 %mAU/g/h).

Substrate kinetics revealed that $q_{glucose}$ was highest toward the end of the batch phase, after which it began to decrease from 0.137±00 g/g/h to 0.072±0.00 g/g/h. A similar trend was observed for $q_{o2}$. $q_{O2}$ was 2.06±0.00 mmol/g/h at 13 h, after which it

gradually decreased to 1.23±0.03 mmol/g/h at 63.5 h. Notably, the biomass-specific consumption rate for EGN2 was much lower than that of AMD. This indicates differences in the physiology of the strains, which might result from strain modification involved with MEP enzymes and downstream enzymes converting FPP to EGN. Based on the information provided by Manus Bio Inc. (Cambridge, USA), it was understood that strain screening in 96DWP resulted in the approximately 3.5% lower performance of the EGN2 strain in comparison with AMD. Thus, scaling up to a higher volume resulted in the improved performance of the EGN2 strain since it exhibited the enhanced expression of the MEP pathway in comparison to the AMD strain. Adequate oxygen provision remains a prominent challenge for aerobic growth in shaken micro-well plate (MWP) fermentations (Maier & Büchs, 2001). By moving from MWP to small-scale bioreactor cultivation, we were able to achieve better oxygen transfer capacity, while EGN2 seemed more sensitive to it than AMD. Regarding the operability of the process, we experienced no foaming issues during the Series H study.

Furthermore, there was a 2-fold improvement in performance index values for the process when compared to the total titer in Series C. The highest specific productivity for the process increased by 45%. Notably, the total phosphate consumed during the process was approximately 12.92 g, of which 5.25 g should have been spent for biomass production. Thus, nearly 5.65 g/L of phosphate could be accounted for MEcPP, which represents a 75% reduction in phosphate drainage to MEcPP when compared to EGN1. Theoretically, only 59.46% of MEcPP is formed in the process while producing close to 290 %AU/L of EGN. This equates to 0.20% of MEcPP per %AU/L of EGN for EGN2 as opposed to 1.31% for EGN1 strain. It was reported that overexpression of *ispG* significantly reduced the efflux of MEcPP, resulting in a 4.5-fold increase in lycopene production (Zhou et al., 2012). Given the strain modification for EGN2, our findings highlight the impact of *ispG* on MEcPP efflux.

### 4.5.3 Mass transfer enhancement at the large scale

Moving toward the end of the thesis, we wanted to address scale-up issues that we faced with EGN1. Hence, we installed a fed-batch cultivation at 300L. As discussed in Section 4.1, we were still dealing with scale-up issues for EGN production in terms of mass transfer limitation, byproduct formation, and cell lysis. Notably, we addressed these issues in the Series I study, where cultivation was performed with the EGN2 strain. In Series D, we were unable to scale up the OTR capacity of the Dasgip system (2L) to 300L. The overall mass transfer rate is not easily measurable because different phenomena are simultaneously occurring during cultivation and their relevance changes with the scale. OTR is influenced by parameters including the physical properties of gas and liquid, operational conditions, the geometric parameters of the bioreactor, and the presence of biomass (Garcia-Ochoa & Gomez, 2009). According to Van't Riet, 1979, the $k_La$ for ionic solutions is more dependent on volumetric power input (P/V) than those for pure water. For our last scale-up study of Series D, power input per unit volume was calculated based on the power number reported for Rushton impellers. Our bioreactor was equipped with two Rushton impellers with a power number $N_p$ of 6, while 2L Dasgip bioreactor units have three Rushton impellers with a $N_p$ of 6. Hence, the maximum P/V for 300L was 3.8 watts/L for Series D, compared to 26.6 watts/L for 2L Dasgip bioreactor. Therefore, the reduction in $k_La$ by power input alone was approximately 75%. To address this issue, we targeted both aeration and agitation. Aeration was increased to 0.3 vvm and agitation was increased to 500 RPM, which was the highest possible stirring rate possible for the bioreactor.

Exponential feeding was put in place using the feed-4 solution (see Section 3.2.6) when glucose concentration was approximately 1 g/L to avoid any form of starvation. The set point for $p_{O2}$ was lowered to 30% to account for the increased solubility of oxygen with a backpressure of 1.2 psi in the bioreactor. Dissolved oxygen levels were controlled using agitation speeds starting from 75 RPM and capped at 500 RPM.

Exponential feeding was switched to a predictive feeding strategy at 24 h to achieve glucose limitation. The set point for $p_{O2}$ was lowered to 15% over a period of 6 h

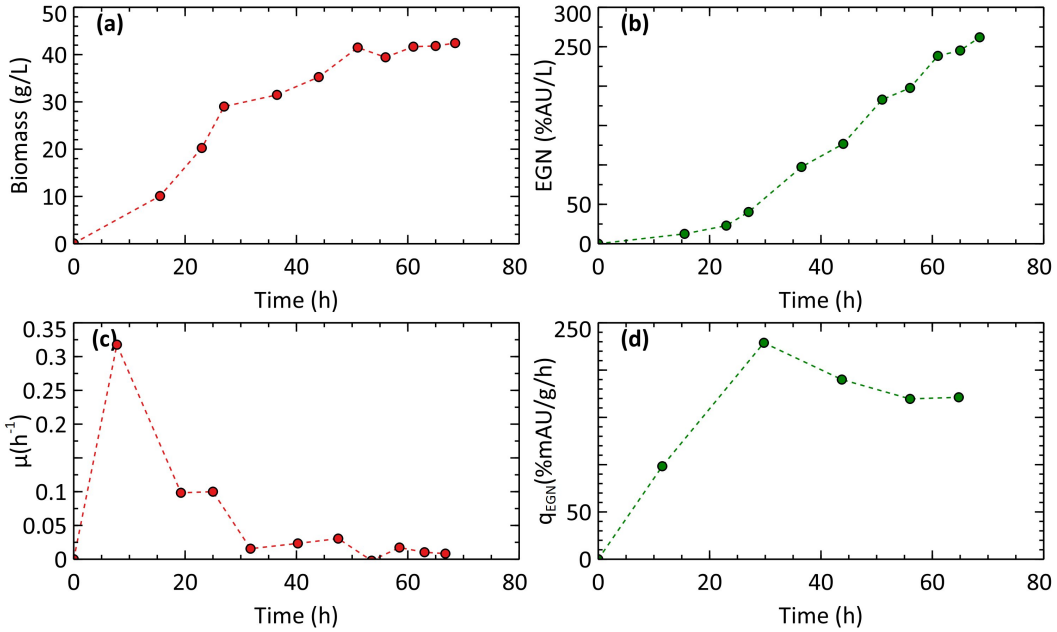



Figure 4.29: Cell density and the production of EGN for Series I. a: Biomass concentration in g/L; Top b: EGN titer in %AU/L; c: Biomass-specific productivity, $q_{EGN}$ in %mAU/g/h; d: Specific growth rate for EGN2, $\mu/$ in $h^{-1}$ during the time course of cultivation.

starting from 24 h. The process was run for 71.5 h.

The highest biomass achieved during cultivation was 39.88 g/L at 65 h, after which it decreased due to a dilution effect (Figure 4.29). EGN production was sustained throughout cultivation to achieve a titer of 243.71 %AU/L after 68.5 h. The specific growth rate and productivity profile matching the 2L cultivation was achieved with productivity remaining above 170 %mAU/g/h. The specific growth rate obtained was as high as 0.32 $h^{-1}$ at 7.75 h while still metabolizing glucose from the batch media, after which it followed a downward trend. Biomass-specific glucose and oxygen kinetics exhibited slightly higher values for Series I when compared to Series H. The lowest $q_{O2}$ for the process was 1.76 mmol/g/g with a corresponding $q_{glucose}$ of 0.058 g/g/h.

The dissolved oxygen level was maintained at 15% during most of the fed-batch. As we increased the final agitation to 500 RPM, the P/V jumped to 17.8 watts/L. Although it was 33% lower than the power input for 2L Dasgip bioreactor, our superficial gas velocity increased by 50% with an increase in aeration to 0.3 vvm. Thus, the increase in $k_La$ based on van't Riet's correlation was 3.27-fold higher than in Series D, amounting to a 1.1-fold higher mass transfer coefficient for Dasgip bioreactors. Upon

scaling up by a ratio of 1:150, the highest possible titer we could achieve in 300L was 82% of the Dasgip system. However, due to heterogeneity, we still need to address the issue of cell oscillation between the nutrient excess and nutrient starvation zones. However, this is beyond the scope of the present study. Despite this, we could maintain the integrity of fermentation broth with reduced foaming.

In general, it is impossible to scale up a bioprocess while keeping all conditions in the optimal values and it remains necessary to choose which variable is the most important. In 2004, Junker (2004) evaluated the different scale-up methods and approaches applied to *E. coli* and yeast processes to find that scale-up from a 30 L laboratory scale typically underpredicted parameters of industrial production, while 280 L scale-up parameters were most similar to the production scale. Thus, by scaling up the process to a 300L volume, we could safely say that our process in a production fermenter should at least be able to reach 82% of the highest titer attained by the EGN2 strain. However, scale-up methodologies never seem to achieve a perfect combination. Thus, novel protocols need to be developed for each process depending on its behavioral traits. For example, when the mass transfer coefficient $k_La$ is relatively high, no influence of oxygen uptake by microorganisms can be detected because the enhancement is very small. However, this situation may change in the case of cultures with high oxygen uptake rates or with high viscous broth, where a decrease in $k_La$ with the activity of microorganisms enhances oxygen mass transfer and $k_La$ increases significantly when compared to physical absorption with no biological influence. For the fermentation of xanthan gum, scale-up had to consider the significant influence of broth viscosity on the mass transfer rate during the production process (Garcıa-Ochoa et al., 2000). Since we conducted a biphasic cultivation, we were also dealing with unknowns related to the impact of oil on the mass transfer of the system. Apart from mass transfer, we need to understand the impact of media and its components at different scales. Thus, there is a lot of scope in the development of scale-up methodologies for terpenoid production with our MEP platform.

# Chapter 5

# Concluding Remarks

## 5.1 Challenges in MEP process development manifested by EGN

The first part of this thesis consisted of developing baseline cultivation conditions for terpenoid production via the MEP pathway in *E. coli* showcasing EGN. The EGN1 strain was engineered with an overexpression of the MEP pathway and the heterologous expression of genes producing the terpenoid EGN. It was observed that the integrity of the process was affected, resulting in breakdown after 35 h. This was evident based on the titer dilution and decreasing biomass due to cell lysis. Coincidentally, lactate production was observed during the phosphate starvation phase. The formation of lactate in *E. coli* could occur via lactate dehydrogenase and the methylglyoxal pathway. With *ldhA* knocked out of the EGN1 strain, the only other possibility was the methylglyoxal pathway, which is known to be active under phosphate-limiting conditions (Chakraborty et al., 2014). This phase was also accompanied by an increase in the specific glucose uptake rate. In 2014, Schuhmacher et al. (2014) demonstrated that phosphate limitation can induce an increased glucose uptake rate as a counteraction for ATP limitation. Since all of the observations pointed toward phosphate starvation as the primary offender for breakdown, we introduced phosphate feeding in the Series C study. A remarkable improvement in process performance was observed,

with titers increasing by 65% and the integrity of the process extending beyond 60 h. *E. coli* biomass only accounts for a maximum of 3% of phosphorus, while we were supplying 5-fold of the required amount. This highlighted the use of phosphate by the MEP pathway, as observed by Banerjee & Sharkey (2014). While our product does not include phosphorus, MEcPP—which is effluxed into the extracellular environment by MEP-overexpressing strains—accounts for 68% of the molar mass (Zhou et al., 2012). Thus, an imbalance of the MEP pathway modules will invariably result in excess phosphate consumption. Although phosphate feeding improved the titer, strain engineering efforts were required to further improve the flux to reach its potential. This was achieved by the EGN2 strain, which had an overexpression of *ispG* and *ispH*, thus resulting in a 2-fold improvement of EGN titer with a leaner process.

## 5.2 Installation of an appropriate glucose and oxygen supply for producing terpenoids in recombinant *E. coli* showcasing amorpha-4,11-diene

In this part of the thesis, the AMD strain was utilized to study the impact of process parameters on the intricate regulation of the MEP pathway. Showcasing AMD4,11 production, this study unraveled the complex metabolic interplay of glucose feeding, oxygen uptake, pyruvate, NADH, ATP, and NADPH supply. Optimum AMD4,11 production could be installed by setting the key process parameters in a balanced way to avoid the drainage of precursors and cofactors. Although shown for AMD4,11 only, the findings are likely to be applied successfully for other MEP-derived terpenoids because the work addressed the fundamental problems of MEP-based terpenoid production. We established 4-fold study with two different systems, namely the monophasic and biphasic systems. The biphasic system, which reduced the active volume by 10% due to the addition of an organic overlayer, increased the substrate availability for cells. It also increased the oxygen demand, thereby increasing the specific oxygen uptake rate. FBA and transcriptome studies importing the experimental observations

revealed higher flux through ATP synthesis reactions, which resulted in NADPH limitation. NADPH is critical to multiple reactions in the MEP pathway (Yang et al., 2016). More importantly, the DXS enzyme can cause DOXP accumulation if limited. This notion was supported by an intracellular metabolomics study indicating the accumulation of DOX (a surrogate for DOXP) for the biphasic system along with the accumulation of a pool for ATP. These issues were resolved with the installation of a 10% lower glucose feed rate, which resulted in a 6.7-fold improvement of the titer and reduction in the by-product accumulation by 67%.

The MEP pathway intrinsically offers well-equilibrated electron balances and the highest glucose-to-product yields to produce terpenoids. Accordingly, these process settings may be qualified as key prerequisites to unlocking the full potential of the MEP pathway. A window of optimum could be implemented to improve the flux through the MEP pathway.

## 5.3 Effect of yeast extract on MEP flux

We concluded that pyruvate and NADPH govern the economics of MEP pathway flux. To further boost the substrate yield, we supplemented the culture with a complex substrate in the form of YE. This is an excellent source of amino acids, peptides, vitamins, and minerals. When amino acids act as a source of carbon and nitrogen, they reduce the activity of the citrate cycle and enhance the pool of pyruvate. Feeding YE reduced the specific oxygen uptake rate and improved the titer by 72%. Although past studies have noted the accumulation of acetic acid, we observed the lower consumption of pH controlling base with YE feeding, which serves as a proxy for organic acid formation. This points toward the efficient use of pyruvate pools for the production of terpenoids via the MEP pathway. Enticing as it may seem, YE also adds to the cost of media development. Therefore, it was important to filter out the components of YE that seem to have a positive impact on the MEP flux. We examined the kinetics of amino acids throughout fermentation to outline the prioritized components. It was observed that the consumption of low-cost amino acids (e.g., aspartate, proline, serine,

arginine, threonine, and phenylalanine) was highest during the growth phase of the batch process, thereby indicating their abundance in the proteomes. During the latter part of the fed-batch, when the specific productivity was higher for YE-fed units, the yield of amino acids on biomass was higher for all amino acids except arginine, glycine, histidine, serine, and tyrosine, which indicates the catabolism of amino acids. Notably, this scenario would increase pyruvate availability. It was concluded via FBA simulation that aspartate, glutamate, serine, glycine, threonine, arginine, alanine, and proline have positive impacts on the flux of the MEP pathway, which correlated well with the experimental observations. However, further studies must be performed with the selected amino acids to prove this hypothesis. With this study, we could further solidify the claim that process development for the MEP pathway-based process can be disentangled by focusing on pyruvate and NADPH.

## 5.4 Process transferability

All of the findings of this thesis were particular to the chassis of the strain and the MEP pathway. As a result, they could easily be transferred to the EGN producer. We performed a follow-up study with the EGN2 strain that was similar to the study involving AMD strain, but with the downstream pathway expressed for EGN production. While we could successfully transfer the process, the EGN titer was increased by 12% with based on IPP production. Downstream pathways can pose a challenge for terpenoid production since they are more oxygen-sensitive for the EGN2 strain when compared to the AMD strain. Scaling up the EGN2 process from a microtiter plate resulted in performance improvement when compared to its counterpart, AMD. Thus, the scale-up criteria for terpenoid production should be more relevant to the mass transfer of oxygen than anything else. Notably, we followed this lead to scale up the process. Since we aimed to match the theoretical mass transfer coefficient from the 2L scale, we increased the power input per unit volume and superficial gas velocity for air. By tightening all of the corners, we successfully scaled up the performance with a titer loss of only 18%. This could have implications for the heterogeneity of the

bioreactor.

# References


Aiba, S., Humphrey, A. E., & Millis, N. F. (1973). *Biochemical engineering* (No. TP 156. F4. A32 1973).

Ajikumar, P. K., Xiao, W.-H., Tyo, K. E., Wang, Y., Simeon, F., Leonard, E., ... Stephanopoulos, G. (2010). Isoprenoid pathway optimization for taxol precursor overproduction in escherichia coli. *Science*, *330*(6000), 70–74.

Akashi, H., & Gojobori, T. (2002). Metabolic efficiency and amino acid composition in the proteomes of escherichia coli and bacillus subtilis. *Proceedings of the National Academy of Sciences*, *99*(6), 3695–3700.

Anders, S., Pyl, P. T., & Huber, W. (2015). Htseq—a python framework to work with high-throughput sequencing data. *Bioinformatics*, *31*(2), 166–169.

Asenjo, J. A. (1994). *Bioreactor system design*. CRC Press.

Banerjee, A., & Sharkey, T. (2014). Methylerythritol 4-phosphate (mep) pathway metabolic regulation. *Natural product reports*, *31*(8), 1043–1055.

Bauer, S., & Ziv, E. (1976). Dense growth of aerobic bacteria in a bench-scale fermentor. *Biotechnology and bioengineering*, *18*(1), 81–94.

Beekwilder, J., van Houwelingen, A., Cankar, K., van Dijk, A. D., de Jong, R. M., Stoopen, G., ... Bosch, D. (2014). Valencene synthase from the heartwood of n ootka cypress (c allitropsis nootkatensis) for biotechnological production of valencene. *Plant biotechnology journal*, *12*(2), 174–182.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, *57*(1), 289–300.

Blount, Z. D. (2015). The natural history of model organisms: The unexhausted potential of e. coli. *Elife*, *4*, e05826.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120.

Buchholz, J., Graf, M., Blombach, B., & Takors, R. (2014, 09). Improving the carbon balance of fermentations by total carbon analyses. *Biochemical Engineering Journal*. doi: 10.1016/j.bej.2014.06.007

Burk, M. J., & Van Dien, S. (2016). Biotechnology for chemical production: challenges and opportunities. *Trends in biotechnology*, *34*(3), 187–190.

Bylund, F., Castan, A., Mikkola, R., Veide, A., & Larsson, G. (2000). Influence of scale-up on the quality of recombinant human growth hormone. *Biotechnology and bioengineering*, *69*(2), 119–128.

Çalik, P., Yilgör, P., Ayhan, P., & Demir, A. S. (2004). Oxygen transfer effects on recombinant benzaldehyde lyase production. *Chemical engineering science*, *59*(22-23), 5075–5083.

Cankar, K., van Houwelingen, A., Goedbloed, M., Renirie, R., de Jong, R. M., Bouwmeester, H., ... Beekwilder, J. (2014). Valencene oxidase cyp706m1 from alaska cedar (callitropsis nootkatensis). *FEBS letters*, *588*(6), 1001–1007.

Chakraborty, S., Karmakar, K., & Chakravortty, D. (2014). Cells producing their own nemesis: understanding methylglyoxal metabolism. *IUBMB life*, *66*(10), 667–678.

Chapman, A. G., Fall, L., & Atkinson, D. E. (1971). Adenylate energy charge in escherichia coli during growth and starvation. *Journal of bacteriology*, *108*(3), 1072–1086.

Chatzivasileiou, A. O., Ward, V., Edgar, S. M., & Stephanopoulos, G. (2019). Two-step pathway for isoprenoid synthesis. *Proceedings of the National Academy of Sciences*, *116*(2), 506–511.

Chen, H., Zhu, C., Zhu, M., Xiong, J., Ma, H., Zhuo, M., & Li, S. (2019). High production of valencene in saccharomyces cerevisiae through metabolic engineering. *Microbial cell factories*, *18*(1), 1–14.

Christ, J. J., & Blank, L. M. (2019). Saccharomyces cerevisiae containing 28% polyphosphate and production of a polyphosphate-rich yeast extract thereof. *FEMS yeast research*, *19*(3), foz011.

Connolly, J. D., & Hill, R. A. (1984). *Dictionary of terpenoids* (No. 03) 547.59 CON).

Daletos, G., Katsimpouras, C., & Stephanopoulos, G. (2020). Novel strategies and platforms for industrial isoprenoid engineering. *Trends in Biotechnology*.

de Kraker, J.-W., Schurink, M., Franssen, M. C., König, W. A., de Groot, A., & Bouwmeester, H. J. (2003). Hydroxylation of sesquiterpenes by enzymes from chicory (cichorium intybus l.) roots. *Tetrahedron*, *59*(3), 409–418.

Del Rio, J. A., Ortuno, A., Garcia-Puig, D., Porras, I., Garcia-Lidon, A., & Sabater, F. (1992). Variations of nootkatone and valencene levels during the development of grapefruit. *Journal of Agricultural and Food Chemistry*, *40*(9), 1488–1490.

De Mey, M., De Maeseneire, S., Soetaert, W., & Vandamme, E. (2007). Minimizing acetate formation in e. coli fermentations. *Journal of industrial microbiology & biotechnology*, *34*(11), 689–700.

Dumont, E., & Delmas, H. (2003). Mass transfer enhancement of gas absorption in oil-in-water systems: a review. *Chemical Engineering and Processing: Process Intensification*, *42*(6), 419–438.

Edwards, J. S., Ibarra, R. U., & Palsson, B. O. (2001). In silico predictions of escherichia coli metabolic capabilities are consistent with experimental data. *Nature biotechnology*, *19*(2), 125–130.

Eisenreich, W., Bacher, A., Arigoni, D., & Rohdich, F. (2004). Biosynthesis of isoprenoids via the non-mevalonate pathway. *Cellular and Molecular Life Sciences CMLS*, *61*(12), 1401–1426.

Eisenreich, W., Schwarz, M., Cartayrade, A., Arigoni, D., Zenk, M. H., & Bacher, A. (1998). The deoxyxylulose phosphate pathway of terpenoid biosynthesis in plants and microorganisms. *Chemistry & biology*, *5*(9), R221–R233.

Eisenstein, E. (1991). Cloning, expression, purification, and characterization of biosynthetic threonine deaminase from escherichia coli. *Journal of Biological Chemistry*, *266*(9), 5801–5807.

Eiteman, M. A., & Altman, E. (2006). Overcoming acetate in escherichia coli recombinant protein fermentations. *Trends in biotechnology*, *24*(11), 530–536.

Erhardt, H., Dempwolff, F., Pfreundschuh, M., Riehle, M., Schäfer, C., Pohl, T., ... Friedrich, T. (2014). Organization of the e scherichia coli aerobic enzyme complexes of oxidative phosphorylation in dynamic domains within the cytoplasmic membrane. *Microbiologyopen*, *3*(3), 316–326.

Erian, A. M., Gibisch, M., & Pflügl, S. (2018). Engineered e. coli w enables efficient 2, 3-butanediol production from glucose and sugar beet molasses using defined minimal medium as economic basis. *Microbial cell factories*, *17*(1), 190.

Escherich, T. (1988). The intestinal bacteria of the neonate and breast-fed infant. *Clinical Infectious Diseases*, *10*(6), 1220–1225.

Eubanks, L. M., & Poulter, C. D. (2003). Rhodobacter capsulatus 1-deoxy-d-xylulose 5-phosphate synthase: steady-state kinetics and substrate binding. *Biochemistry*, *42*(4), 1140–1149.

Farmer, W. R., & Liao, J. C. (2001). Precursor balancing for metabolic engineering of lycopene production in escherichia coli. *Biotechnology progress*, *17*(1), 57–61.

Fenton, D., Lai, H., Lu, H., Mann, M., & Tsai, L. (1997, February 4). *Control of norleucine incorporation into recombinant proteins.* Google Patents. (US Patent 5,599,690)

Fraatz, M. A., Riemer, S. J., Stöber, R., Kaspera, R., Nimtz, M., Berger, R. G., & Zorn, H. (2009). A novel oxygenase from pleurotus sapidus transforms valencene to nootkatone. *Journal of Molecular Catalysis B: Enzymatic*, *61*(3-4), 202–207.

Garcia-Ochoa, F., & Gomez, E. (2009). Bioreactor scale-up and oxygen transfer rate in microbial processes: an overview. *Biotechnology advances*, *27*(2), 153–176.

Garcıa-Ochoa, F., Santos, V., Casas, J., & Gómez, E. (2000). Xanthan gum: production, recovery, and properties. *Biotechnology advances*, *18*(7), 549–579.

Gavira, C., Höfer, R., Lesot, A., Lambert, F., Zucca, J., & Werck-Reichhart, D. (2013). Challenges and pitfalls of p450-dependent (+)-valencene bioconversion by saccharomyces cerevisiae. *Metabolic engineering*, *18*, 25–35.

Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., ... others (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome biology*, *5*(10), R80.

Giner, J.-L., & Jaun, B. (1998). Biosynthesis of isoprenoids in escherichia coli: Retention of the methyl h-atoms of 1-deoxy-d-xylulose. *Tetrahedron letters*, *39*(44), 8021–8022.

Girhard, M., Machida, K., Itoh, M., Schmid, R. D., Arisawa, A., & Urlacher, V. B. (2009). Regioselective biooxidation of (+)-valencene by recombinant e. coli expressing cyp109b1 from bacillus subtilis in a two-liquid-phase system. *Microbial cell factories*, *8*(1), 1–12.

Goldstein, J. L., & Brown, M. S. (1990). Regulation of the mevalonate pathway. *Nature*, *343*(6257), 425–430.

Goodwin, T. (1958). Studies in carotenogenesis. 25. the incorporation of 14co2,[2-14c] acetate and [2-14c] mevalonate into $\beta$-carotene by illuminated etiolated maize seedlings. *Biochemical Journal*, *70*(4), 612–617.

Gordienko, E. N., Kazanov, M. D., & Gelfand, M. S. (2013). Evolution of pan-genomes of escherichia coli, shigella spp., and salmonella enterica. *Journal of bacteriology*, *195*(12), 2786–2792.

Gronemeyer, P., Ditz, R., & Strube, J. (2014). Trends in upstream and downstream process development for antibody manufacturing. *Bioengineering*, *1*(4), 188–212.

Gruchattka, E., Hädicke, O., Klamt, S., Schütz, V., & Kayser, O. (2013). In silico profiling of escherichia coli and saccharomyces cerevisiae as terpenoid factories. *Microbial cell factories*, *12*(1), 84.

Hahn, F. M., Hurlburt, A. P., & Poulter, C. D. (1999). Escherichia coli open reading frame 696 is idi, a nonessential gene encoding isopentenyl diphosphate isomerase. *Journal of bacteriology*, *181*(15), 4499–4504.

Hale, V., Keasling, J. D., Renninger, N., & Diagana, T. T. (2007). Microbially derived artemisinin: a biotechnology solution to the global problem of access to affordable antimalarial drugs. *The American journal of tropical medicine and hygiene*, *77*(6_Suppl), 198–202.

Han, Q., & Eiteman, M. A. (2019). Acetate formation during recombinant protein production in escherichia coli k-12 with an elevated nad (h) pool. *Engineering in Life Sciences*, *19*(11), 770–780.

Haring, H. G., Rijkens, F., Boelens, H., & Van der Gen, A. (1972). Olfactory studies on enantiomeric eremophilane sesquiterpenoids. *Journal of Agricultural and Food Chemistry*, *20*(5), 1018–1021.

Harker, M., & Bramley, P. (1999). Expression of prokaryotic 1-deoxy-d-xylulose-5-phosphatases in escherichia coli increases carotenoid and ubiquinone biosynthesis. *FEBS letters*, *448*(1), 115–119.

Herbst, H., Schumpe, A., & Deckwer, W.-D. (1992). Xanthan production in stirred tank fermenters: Oxygen transfer and scale-up. *Chemical Engineering & Technology: Industrial Chemistry-Plant Equipment-Process Engineering-Biotechnology*, *15*(6), 425–434.

Herz, S., Wungsintaweekul, J., Schuhr, C. A., Hecht, S., Lüttgen, H., Sagner, S., ... others (2000). Biosynthesis of terpenoids: Ygbb protein converts 4-diphosphocytidyl-2c-methyl-d-erythritol 2-phosphate to 2c-methyl-d-erythritol 2, 4-cyclodiphosphate. *Proceedings of the National Academy of Sciences*, *97*(6), 2486–2490.

Hill, R. E., Sayer, B. G., & Spenser, I. D. (1989). Biosynthesis of vitamin b6: incorporation of d-1-deoxyxylulose. *Journal of the American Chemical Society*, *111*(5), 1916–1917.

Hunter, G., & Brogden Jr, W. (1965). Analysis of the terpene and sesquiterpene hydrocarbons in some citrus oils. *Journal of Food Science*, *30*(3), 383–387.

ICCA, & Oxford Economics. (2019). The Global Chemical Industry: Catalyzing Growth and Addressing Our World's Sustainability Challenges. (March), 1–29.

Imlay, J. A. (2013). The molecular mechanisms and physiological consequences of oxidative stress: lessons from a model bacterium. *Nature Reviews Microbiology*, *11*(7), 443–454.

Iuchi, S., & Lin, E. (1993). Adaptation of escherichia coli to redox environments by gene expression. *Molecular microbiology*, *9*(1), 9–15.

Junker, B. H. (2004). Scale-up methodologies for escherichia coli and yeast fermentation processes. *Journal of bioscience and bioengineering*, *97*(6), 347–364.

Kaleta, C., Schäuble, S., Rinas, U., & Schuster, S. (2013). Metabolic costs of amino acid and protein production in escherichia coli. *Biotechnology journal*, *8*(9), 1105–1114.

Kaneda, K., Kuzuyama, T., Takagi, M., Hayakawa, Y., & Seto, H. (2001). An unusual isopentenyl diphosphate isomerase found in the mevalonate pathway gene cluster from streptomyces sp. strain cl190. *Proceedings of the National Academy of Sciences*, *98*(3), 932–937.

Kim, S.-W., & Keasling, J. (2001). Metabolic engineering of the nonmevalonate isopentenyl diphosphate synthesis pathway in escherichia coli enhances lycopene production. *Biotechnology and bioengineering*, *72*(4), 408–415.

Kirby, J., & Keasling, J. D. (2009). Biosynthesis of plant isoprenoids: perspectives for microbial engineering. *Annual review of plant biology*, *60*, 335–355.

Kitano, H. (2002). Systems biology: a brief overview. *science*, *295*(5560), 1662–1664.

Krishna, S., Bustamante, L., Haynes, R. K., & Staines, H. M. (2008). Artemisinins: their growing importance in medicine. *Trends in pharmacological sciences*, *29*(10), 520–527.

Kumaran, A. P., Ryan, L., Donald, J., Tseng, H.-c., Santos, C., & Philippe, R. (2019, November 19). *Metabolic engineering for microbial production of terpenoid products.* Google Patents. (US Patent 10,480,015)

Kuzuyama, T., Takagi, M., Takahashi, S., & Seto, H. (2000). Cloning and characterization of 1-deoxy-d-xylulose 5-phosphate synthase fromstreptomyces sp. strain cl190, which uses both the mevalonate and nonmevalonate pathways for isopentenyl diphosphate biosynthesis. *Journal of Bacteriology*, *182*(4), 891–897.

Kvint, K., Nachin, L., Diez, A., & Nyström, T. (2003). The bacterial universal stress protein: function and regulation. *Current opinion in microbiology*, *6*(2), 140–145.

Lange, B. M., Rujan, T., Martin, W., & Croteau, R. (2000). Isoprenoid biosynthesis: the evolution of two ancient and distinct pathways across genomes. *Proceedings of the National Academy of Sciences*, *97*(24), 13172–13177.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with bowtie 2. *Nature methods*, *9*(4), 357.

Laurent, P., Braekman, J.-C., Daloze, D., & Pasteels, J. (2003). Biosynthesis of defensive compounds from beetles and ants. *European Journal of Organic Chemistry*, *2003*(15), 2733–2743.

Leonard, E., Ajikumar, P. K., Thayer, K., Xiao, W.-H., Mo, J. D., Tidor, B., ... Prather, K. L. (2010). Combining metabolic and protein engineering of a terpenoid biosynthetic pathway for overproduction and selectivity control. *Proceedings of the National Academy of Sciences*, *107*(31), 13654–13659.

Lindroth, P., & Mopper, K. (1979, 09). High performance liquid chromatographic determination of subpicomole amounts of amino acids by precolumn fluorescence derivatization with o-phthaldialdehyde. *Analytical Chemistry - ANAL CHEM*, *51*. doi: 10.1021/ac50047a019

Little, H. N., & Bloch, K. (1950). Studies on the utilization of acetic acid for the biological synthesis of cholesterol. *Journal of Biological Chemistry*, *183*(1), 33–46.

Liu, H., Sun, Y., Ramos, K. R. M., Nisola, G. M., Valdehuesa, K. N. G., Lee, W.-K., ... Chung, W.-J. (2013). Combination of entner-doudoroff pathway with mep increases isoprene production in engineered escherichia coli. *PLoS One*, *8*(12), e83290.

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for rna-seq data with deseq2. *Genome biology*, *15*(12), 550.

Lu, C., Albano, C. R., Bentley, W. E., & Rao, G. (2005). Quantitative and kinetic study of oxidative stress regulons using green fluorescent protein. *Biotechnology and bioengineering*, *89*(5), 574–587.

Luli, G. W., & Strohl, W. R. (1990). Comparison of growth, acetate production, and acetate inhibition of escherichia coli strains in batch and fed-batch fermentations. *Applied and environmental microbiology*, *56*(4), 1004–1011.

Luo, W., & Brouwer, C. (2013). Pathview: an r/bioconductor package for pathway-based data integration and visualization. *Bioinformatics*, *29*(14), 1830–1831.

Luo, W., Friedman, M. S., Shedden, K., Hankenson, K. D., & Woolf, P. J. (2009). Gage: generally applicable gene set enrichment for pathway analysis. *BMC bioinformatics*, *10*(1), 161.

Lütke-Brinkhaus, F., & Kleinig, H. (1987). Formation of isopentenyl diphosphate via mevalonate does not occur within etioplasts and etiochloroplasts of mustard (sinapis alba l.) seedlings. *Planta*, *171*(3), 406–411.

Lüttgen, H., Rohdich, F., Herz, S., Wungsintaweekul, J., Hecht, S., Schuhr, C. A., ... others (2000). Biosynthesis of terpenoids: Ychb protein of escherichia coli phosphorylates the 2-hydroxy group of 4-diphosphocytidyl-2c-methyl-d-erythritol. *Proceedings of the National Academy of Sciences*, *97*(3), 1062–1067.

Lyubetskaya, A., Rubanov, L., & Gelfand, M. (2006). Use of the flux model of amino acid metabolism of escherichia coli. *Biochemistry (moscow)*, *71*(11), 1256–1260.

Maier, U., & Büchs, J. (2001). Characterisation of the gas–liquid mass transfer in shaking bioreactors. *Biochemical Engineering Journal*, *7*(2), 99–106.

Margaritis, A. (1978). Biotechology review: mixing, mass transfer, and scale-up of polysaccharide fermentations. *Biotechnol. Bioeng.*, *20*, 977–993.

Matthäus, F., Ketelhot, M., Gatter, M., & Barth, G. (2014). Production of lycopene in the non-carotenoid-producing yeast yarrowia lipolytica. *Applied and Environmental Microbiology*, *80*(5), 1660–1669.

Meadows, A. L., Hawkins, K. M., Tsegaye, Y., Antipov, E., Kim, Y., Raetz, L., ... others (2016). Rewriting yeast central carbon metabolism for industrial isoprenoid production. *Nature*, *537*(7622), 694–697.

Mi, J., Becher, D., Lubuta, P., Dany, S., Tusch, K., Schewe, H., ... Schrader, J. (2014). De novo production of the monoterpenoid geranic acid by metabolically engineered pseudomonas putida. *Microbial cell factories*, *13*(1), 170.

Mikami, Y., Yoneda, H., Tatsukami, Y., Aoki, W., & Ueda, M. (2017). Ammonia production from amino acid-based biomass-like sources by engineered escherichia coli. *AMB Express*, *7*(1), 83.

Moser, S., & Pichler, H. (2019). Identifying and engineering the ideal microbial terpenoid production host. *Applied microbiology and biotechnology*, *103*(14), 5501–5516.

Nakamura, C. E., & Whited, G. M. (2003). Metabolic engineering for the microbial production of 1, 3-propanediol. *Current opinion in biotechnology*, *14*(5), 454–459.

Noor, R. (2015). Mechanism to control the cell lysis and the cell survival strategy in stationary phase under heat stress. *Springerplus*, *4*(1), 1–9.

Orth, J. D., Thiele, I., & Palsson, B. Ø. (2010). What is flux balance analysis? *Nature biotechnology*, *28*(3), 245–248.

Paddon, C. J., & Keasling, J. D. (2014). Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development. *Nature Reviews Microbiology*, *12*(5), 355–367.

Paddon, C. J., Westfall, P. J., Pitera, D. J., Benjamin, K., Fisher, K., McPhee, D., ... others (2013). High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature*, *496*(7446), 528–532.

Patil, V., Santos, C. N. S., Ajikumar, P. K., Sarria, S., & Takors, R. (2020). Balancing glucose and oxygen uptake rates to enable high amorpha-4, 11-diene production in

escherichia coli via the methylerythritol phosphate (mep) pathway. *Biotechnology and Bioengineering*.

Peplow, M. (2016). Synthetic biology's first malaria drug meets market resistance. *Nature News*, *530*(7591), 389.

Proust, L., Sourabié, A., Pedersen, M., Besançon, I., Haudebourg, E., Monnet, V., & Juillard, V. (2019). Insights into the complexity of yeast extract peptides and their utilization by streptococcus thermophilus. *Frontiers in microbiology*, *10*, 906.

Quin, M. B., Flynn, C. M., & Schmidt-Dannert, C. (2014). Traversing the fungal terpenome. *Natural product reports*, *31*(10), 1449–1473.

Ramawat, K. G., & Mérillon, J.-M. (2013). *Natural products: phytochemistry, botany and metabolism of alkaloids, phenolics and terpenes*. Springer.

Reiling, H., Laurila, H., & Fiechter, A. (1985). Mass culture of escherichia coli: Medium development for low and high density cultivation of escherichia coli b/r in minimal and complex media. *Journal of Biotechnology*, *2*(3-4), 191–206.

Renault, H., Bassard, J.-E., Hamberger, B., & Werck-Reichhart, D. (2014). Cytochrome p450-mediated metabolic engineering: current progress and future challenges. *Current opinion in plant biology*, *19*, 27–34.

Riesenberg, D., Schulz, V., Knorre, W., Pohl, H.-D., Korz, D., Sanders, E., ... Deckwer, W.-D. (1991). High cell density cultivation of escherichia coli at controlled specific growth rate. *Journal of biotechnology*, *20*(1), 17–27.

Rogers, K., & Kadner, R. (n.d.). Facts about e. coli: Dimensions, as discussed in bacteria: Diversity of structure of bacteria. *Encyclopedia Britannica Online*.

Rohdich, F., Hecht, S., Gärtner, K., Adam, P., Krieger, C., Amslinger, S., ... Eisenreich, W. (2002). Studies on the nonmevalonate terpene biosynthetic pathway: metabolic role of isph (lytb) protein. *Proceedings of the National Academy of Sciences*, *99*(3), 1158–1163.

Rohdich, F., Wungsintaweekul, J., Fellermeier, M., Sagner, S., Herz, S., Kis, K., ... Zenk, M. H. (1999). Cytidine 5-triphosphate-dependent biosynthesis of isoprenoids: Ygbp protein of escherichia coli catalyzes the formation of 4-diphosphocytidyl-2-c-methylerythritol. *Proceedings of the National Academy of Sciences*, *96*(21), 11758–11763.

Rohmer, M., Knani, M., Simonin, P., Sutter, B., & Sahm, H. (1993). Isoprenoid biosynthesis in bacteria: a novel pathway for the early steps leading to isopentenyl diphosphate. *Biochemical Journal*, *295*(2), 517–524.

Romero, I. G., Pai, A. A., Tung, J., & Gilad, Y. (2014). Rna-seq: impact of rna degradation on transcript quantification. *BMC biology*, *12*(1), 1–13.

Rothen, S., Sauer, M., Sonnleitner, B., & Witholt, B. (1998). Growth characteristics of escherichia colihb101 [pgec47] on defined medium. *Biotechnology and bioengineering*, *58*(1), 92–100.

Sacchettini, J. C., & Poulter, C. D. (1997). Creating isoprenoid diversity. *Science*, *277*(5333), 1788–1789.

Salgado, H., Gama-Castro, S., Martínez-Antonio, A., Díaz-Peredo, E., Sánchez-Solano, F., Peralta-Gil, M., ... others (2004). Regulondb (version 4.0): transcriptional regulation, operon organization and growth conditions in escherichia coli k-12. *Nucleic Acids Research*, *32*(suppl_1), D303–D306.

Schellenberger, J., Que, R., Fleming, R. M., Thiele, I., Orth, J. D., Feist, A. M., ... others (2011). Quantitative prediction of cellular metabolism with constraint-based models: the cobra toolbox v2. 0. *Nature protocols*, *6*(9), 1290.

Schmidt, F. (2005). Optimization and scale up of industrial fermentation processes. *Applied microbiology and biotechnology*, *68*(4), 425–435.

Schörken, U., & Sprenger, G. A. (1998). Thiamin-dependent enzymes as catalysts in chemoenzymatic syntheses. *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology*, *1385*(2), 229–243.

Schuhmacher, T., Löffler, M., Hurler, T., & Takors, R. (2014). Phosphate limited fed-batch processes: impact on carbon usage and energy metabolism in escherichia coli. *Journal of Biotechnology*, *190*, 96–104.

Schürmann, M., Schürmann, M., & Sprenger, G. A. (2002). Fructose 6-phosphate aldolase and 1-deoxy-d-xylulose 5-phosphate synthase from escherichia coli as tools in enzymatic synthesis of 1-deoxysugars. *Journal of Molecular Catalysis B: Enzymatic*, *19*, 247–252.

Seemann, M., Bui, B. T. S., Wolff, M., Tritsch, D., Campos, N., Boronat, A., ... Rohmer, M. (2002). Isoprenoid biosynthesis through the methylerythritol phosphate pathway: The (e)-4-hydroxy-3-methylbut-2-enyl diphosphate synthase (gcpe) is a [4fe–4s] protein. *Angewandte Chemie*, *114*(22), 4513–4515.

Seifert, A., Vomund, S., Grohmann, K., Kriening, S., Urlacher, V. B., Laschat, S., & Pleiss, J. (2009). Rational design of a minimal and highly enriched cyp102a1 mutant library with improved regio-, stereo-, and chemo-selectivity. *Cytochrom P450-Monooxygenasen: Modellierung, Datenbankanalyse und experimentelle Charakterisierung neuer Enzymvarianten*, *597*, 83.

Sharfstein, S. T., & Keasling, J. (1994). Polyphosphate metabolism in escherichia coli. *Annals of the New York Academy of Sciences*, *745*(1), 77–91.

Shiloach, J., & Fass, R. (2005). Growing e. coli to high cell density—a historical perspective on method development. *Biotechnology advances*, *23*(5), 345–357.

Shimizu, K. (2014). Regulation systems of bacteria such as escherichia coli in response to nutrient limitation and environmental stresses. *Metabolites*, *4*(1), 1–35.

Sigüenza, R., Flores, N., Hernández, G., Martínez, A., Bolivar, F., & Valle, F. (1999). Kinetic characterization in batch and continuous culture of escherichia coli mutants affected in phosphoenolpyruvate metabolism: differences in acetic acid production. *World Journal of Microbiology and Biotechnology*, *15*(5), 587–592.

Sikkema, J., de Bont, J. A., & Poolman, B. (1995). Mechanisms of membrane toxicity of hydrocarbons. *Microbiological reviews*, *59*(2), 201–222.

Soini, J., Ukkonen, K., & Neubauer, P. (2008). High cell density media for escherichia coli are generally designed for aerobic cultivations–consequences for large-scale bioprocesses and shake flask cultures. *Microbial Cell Factories*, *7*(1), 1–11.

Sowden, R. J., Yasmin, S., Rees, N. H., Bell, S. G., & Wong, L.-L. (2005). Biotransformation of the sesquiterpene (+)-valencene by cytochrome p450 cam and p450 bm-3. *Organic & biomolecular chemistry*, *3*(1), 57–64.

Spaans, S. K., Weusthuis, R. A., Van Der Oost, J., & Kengen, S. W. (2015). Nadph-generating systems in bacteria and archaea. *Frontiers in microbiology*, *6*, 742.

Sprenger, G. A., Schörken, U., Wiegert, T., Grolle, S., De Graaf, A. A., Taylor, S. V., ... Sahm, H. (1997). Identification of a thiamin-dependent synthase in escherichia coli required for the formation of the 1-deoxy-d-xylulose 5-phosphate precursor to isoprenoids, thiamin, and pyridoxol. *Proceedings of the National Academy of Sciences*, *94*(24), 12857–12862.

Suárez, D., Liria, C., & Kilikian, B. (1998). Effect of yeast extract on escherichia coli growth and acetic acid production. *World Journal of Microbiology and Biotechnology*, *14*(3), 331–335.

Suffness, M. (1995). *Taxol: science and applications* (Vol. 22). CRC press.

Takahashi, C. M., Takahashi, D. F., Carvalhal, M. L., & Alterthum, F. (1999). Effects of acetate on the growth and fermentation performance of escherichia coli ko11. *Applied biochemistry and biotechnology*, *81*(3), 193–203.

Takahashi, S., Kuzuyama, T., Watanabe, H., & Seto, H. (1998). A 1-deoxy-d-xylulose 5-phosphate reductoisomerase catalyzing the formation of 2-c-methyl-d-erythritol 4-phosphate in an alternative nonmevalonate pathway for terpenoid biosynthesis. *Proceedings of the National Academy of Sciences*, *95*(17), 9879–9884.

Thompson, B., Kole, M., & Gerson, D. (1985). Control of ammonium concentration in escherichia coli fermentations. *Biotechnology and bioengineering*, *27*(6), 818–824.

Tian, H., Zada, B., Singh, B. H., Wang, C., & Kim, S.-W. (2019). Synthetic biology approaches for the production of isoprenoids in escherichia coli. In *Current developments in biotechnology and bioengineering* (pp. 311–329). Elsevier.

Tsuruta, H., Paddon, C. J., Eng, D., Lenihan, J. R., Horning, T., Anthony, L. C., ... Newman, J. D. (2009). High-level production of amorpha-4, 11-diene, a precursor of the antimalarial agent artemisinin, in escherichia coli. *PLoS One*, *4*(2), e4489.

van Rijsewijk, B. R. H., Kochanowski, K., Heinemann, M., & Sauer, U. (2016). Distinct transcriptional regulation of the two escherichia coli transhydrogenases pntab and udha. *Microbiology*, *162*(9), 1672–1679.

Van't Riet, K. (1979). Review of measuring methods and results in nonviscous gas-liquid mass transfer in stirred vessels. *Industrial & Engineering Chemistry Process Design and Development*, *18*(3), 357–364.

Vasilakou, E., Van Loosdrecht, M. C., & Wahl, S. A. (2020). Escherichia coli metabolism under short-term repetitive substrate dynamics: adaptation and trade-offs. *Microbial Cell Factories*, *19*, 1–19.

Weiss, M., Haufe, J., Carus, M., Brandão, M., Bringezu, S., Hermann, B., & Patel, M. K. (2012). A review of the environmental impacts of biobased materials. *Journal of Industrial Ecology*, *16*, S169–S181.

Welch, R. A., Burland, V., Plunkett, G., Redford, P., Roesch, P., Rasko, D., ... others (2002). Extensive mosaic structure revealed by the complete genome sequence of uropathogenic escherichia coli. *Proceedings of the National Academy of Sciences*, *99*(26), 17020–17024.

Westfall, P. J., Pitera, D. J., Lenihan, J. R., Eng, D., Woolard, F. X., Regentin, R., ... others (2012). Production of amorphadiene in yeast, and its conversion to

dihydroartemisinic acid, precursor to the antimalarial agent artemisinin. *Proceedings of the National Academy of Sciences*, *109*(3), E111–E118.

Wewetzer, S., Kunze, M., Ladner, T., Luchterhand, B., Roth, S., Rahmen, N., ... Büchs, J. (2015). Parallel use of shake flask and microtiter plate online measuring devices (ramos and biolector) reduces the number of experiments in laboratory-scale stirred tank bioreactors. *Journal of biological engineering*, *9*(1), 9.

WHO. (2019). *World Malaria Report. https://www.who.int/malaria*. Retrieved from `https://www.who.int/publications-detail/world-malaria-report-2019`

Wilson III, C. W., & Shaw, P. E. (1978). Synthesis of nootkatone from valencene. *Journal of Agricultural and Food Chemistry*, *26*(6), 1430–1432.

Wong, U., & Cox, R. J. (2007). The chemical mechanism of d-1-deoxyxylulose-5-phosphate reductoisomerase from escherichia coli. *Angewandte Chemie International Edition*, *46*(26), 4926–4929.

World Health Organization. (2018). Status report on artemisinin resistance and ACT efficacy (August 2018). *World Health Organization*(August), 10. Retrieved from `https://www.who.int/malaria/publications/atoz/artemisinin`

Wu, G., Yan, Q., Jones, J. A., Tang, Y. J., Fong, S. S., & Koffas, M. A. (2016). Metabolic burden: cornerstones in synthetic biology and metabolic engineering applications. *Trends in biotechnology*, *34*(8), 652–664.

Yadav, V. G., De Mey, M., Lim, C. G., Ajikumar, P. K., & Stephanopoulos, G. (2012). The future of metabolic engineering and synthetic biology: towards a systematic practice. *Metabolic engineering*, *14*(3), 233–241.

Yamada, Y., Kuzuyama, T., Komatsu, M., Shin-ya, K., Omura, S., Cane, D. E., & Ikeda, H. (2015). Terpene synthases are widely distributed in bacteria. *Proceedings of the National Academy of Sciences*, *112*(3), 857–862.

Yang, C., Gao, X., Jiang, Y., Sun, B., Gao, F., & Yang, S. (2016). Synergy between methylerythritol phosphate pathway and mevalonate pathway for isoprene production in escherichia coli. *Metabolic engineering*, *37*, 79–91.

Yee, L., & Blanch, H. (1992). Recombinant protein expression in high cell density fed-batch cultures of escherichia coli. *Bio/technology*, *10*(12), 1550–1556.

Zampieri, M., Hörl, M., Hotz, F., Müller, N. F., & Sauer, U. (2019). Regulatory mechanisms underlying coordination of amino acid and glucose catabolism in escherichia coli. *Nature communications*, *10*(1), 1–13.

Zhang, J. (2005). A detailed chronological record of project 523 and the discovery and development of qinghaosu (artemisinin) vol. 193 yangcheng evening news publishing company. *China.[Google Scholar]*.

Zhao, L., Chang, W.-c., Xiao, Y., Liu, H.-w., & Liu, P. (2013). Methylerythritol phosphate pathway of isoprenoid biosynthesis. *Annual review of biochemistry*, *82*, 497.

Zhou, K., Zou, R., Stephanopoulos, G., & Too, H.-P. (2012). Metabolite profiling identified methylerythritol cyclodiphosphate efflux as a limiting step in microbial isoprenoid production. *PLoS One*, *7*(11), e47513.

Zhu, J., Sanchez, A., Bennett, G. N., & San, K.-Y. (2011). Manipulating respiratory levels in escherichia coli for aerobic formation of reduced chemical products. *Metabolic engineering*, *13*(6), 704–712.

Zhu, Y., Rinzema, A., Tramper, J., & Bol, J. (1996). Medium design based on stoichiometric analysis of microbial transglutaminase production by streptoverticillium mobaraense. *Biotechnology and bioengineering*, *50*(3), 291–298.

# Author Contribution

This chapter summarizes my (Vikas Patil) contribution to the manuscript that was already published. The contents of the manuscript are supplemented in the Appendix. (Appendix A).

## Manuscript

**Patil, V.**, Santos, C., Ajikumar, P., Sarria S., & Takors R. (2020). Balancing glucose and oxygen uptake rates to enable high amorpha-4,11-diene production in *Escherichia coli* via the methylerythritol phosphate pathway. *Biotechnology and Bioengineering.*

# Declaration of Originality

I declare that the submitted work has been completed by me and that I have not used any other than permitted reference sources or materials. All references and other sources used by me have been appropriately acknowledged in the work.

*San Mateo, CA, USA,*
*13.01.2022*
Place, Date

Vikas Patil
Vikas Patil

# Vikas Patil

Augusta, GA | (706)814-2745
vhpatil91@gmail.com | www.linkedin.com/in/vikas-haridas-patil

## Summary

-Bioprocess engineer with chemical engineering background and an extensive experience in fermentation process development, scale up and technology transfer.
-Highly skilled in statistical analysis, strain characterization and lab management.

## Key Skills

- Technology Transfer
- Scale-up and Scale-down
- Design of Experiments
- Statistical Analysis
- Process Control
- MATLAB
- Excel, JMP, Minitab
- Aspen, COMSOL
- HPLC-DAD, GC-MS/FID/Ion Trap
- Flux Balance Analysis
- Aseptic Techniques
- Downstream Processing
- Metabolic Engineering
- Microbial Physiology
- Omics

## Experience

### Manus Bio, Inc., Augusta, GA

**Senior Process Development Engineer-Fermentation** Feb '21- Present

- Supervise a team of associate scientists in the design and implementation of upstream bioreactor process ranging from 400 µL to 300L vessels.
- Support in commercialization and product development efforts for multiple products in the pipeline.
- Scale up process from 250 mL bioreactors to the pilot plant (300-3000L scale).
- Collaborate with Analytics team to establish a fermentation process with state-of-the-art analytical capabilities, including metabolomics and proteomics.
- Build a line of communication with upstream and downstream teams to develop consistent and robust operations across the company.
- Provide technical expertise to manufacturing team for continuous improvement, scale up and troubleshooting of production operations.
- Develop technical strategies and design scale-down studies to provide a statistical data-driven approach to process optimization and troubleshooting.

**Process Development Engineer II-Fermentation** Sep '19- Feb'21

- Lead process development of a product to reach commercial titer goal, achieving 30-fold improvement in process productivity.
- Worked with a team of scientist to optimize the media, reducing the cost of the cultivation by ~50%.
- Established efficient coordination, communication, and training of team members to increase experimental throughput by ~33%.
- Established a high throughput platform with 96DWP and small-scale bioreactors for strain screening and characterization.
- Developed analytical methods on HPLC-DAD/RID and GC-FID/MS to quantify metabolic intermediates and natural products.

**Research Associate-Bioprocess Development** Sep '15- Apr'16

- Routinely prepared media, set up operation and harvest of bench-scale bioreactors.
- Assisted in design of experiments targeting strain screening and evaluation, media optimization and process development in scale ranging from shake flasks to 10 litre bioreactors.
- Processed and analysed data and reported findings to all the teams to suggest direction to their research.
- Maintained clean and organized lab environment as well as lab notebooks, reports and SOPs.

## Publication

Patil, V, Santos, CNS, Ajikumar, PK, Sarria, S, Takors, R. Balancing glucose and oxygen uptake rates to enable high amorpha-4,11-diene production in Escherichia coli via the methylerythritol phosphate pathway. Biotechnology and Bioengineering. 2021; 118: 1317– 1329. https://doi.org/10.1002/bit.27655

## Education

- Ph.D. in Biochemical Engineering, Institute of Biochemical Engineering, University of Stuttgart, Stuttgart, Germany; Expected 2021.
- M.S. in Chemical Engineering, Arizona State University, Tempe, AZ, USA; 2015
- B. Tech. in Chemical Engineering, Vishwakarma Institute of Technology, Pune, India; 2013

# Appendix A

# Manuscript

## Balancing glucose and oxygen uptake rates to enable high amorpha-4,11-diene production in *Escherichia coli* via the methylerythritol phosphate pathway

Received: 15 September 2020 | Revised: 29 November 2020 | Accepted: 12 December 2020

DOI: 10.1002/bit.27655



ARTICLE

BIOTECHNOLOGY AND BIOENGINEERING WILEY

# Balancing glucose and oxygen uptake rates to enable high amorpha-4,11-diene production in *Escherichia coli* via the methylerythritol phosphate pathway

Vikas Patil[1,2] | Christine N. S. Santos[2] | Parayil K. Ajikumar[2] | Stephen Sarria[2] | Ralf Takors[1]

[1]Institute of Biochemical Engineering, University of Stuttgart, Stuttgart, Germany

[2]Manus Bio, Cambridge, Massachusetts, USA

**Correspondence**
Ralf Takors, Institute of Biochemical Engineering, University of Stuttgart, Allmandring 31, D-70569 Stuttgart, Germany.
Email: ralf.takors@ibvt.uni-stuttgart.de

**Funding information**
Manus Bio, Inc.

## Abstract

Amorpha-4,11-diene (AMD4,11) is a precursor to artemisinin, a potent antimalarial drug that is traditionally extracted from the shrubs of *Artemisia annua*. Despite significant prior efforts to produce artemisinin and its precursors through biotechnology, there remains a dire need for more efficient biosynthetic routes for its production. Here, we describe the optimization of key process conditions for an *Escherichia coli* strain producing AMD4,11 via the native methylerythritol phosphate (MEP) pathway. By studying the interplay between glucose uptake rates and oxygen demand, we were able to identify optimal conditions for increasing carbon flux through the MEP pathway by manipulating the availability of NADPH required for terpenoid production. Installation of an optimal $q_{O2}/q_{glucose}$ led to a 6.7-fold increase in product titers and a 6.5-fold increase in carbon yield.

**KEYWORDS**
amorpha-4,11-diene, FBA, MEP pathway, NADPH, oxygen uptake, terpenoid

## 1 | INTRODUCTION

Malaria is a life-threatening disease caused by parasites within the *Plasmodium* group. In 2018, the World Health Organization (WHO) reported around 228 million cases of malaria, with 0.17% or 405,000 of these cases ultimately ending in death (WHO Report, 2019). Most of these deaths were attributed to infections by the parasite *P. falciparum*, which has unfortunately become resistant to most of the conventionally used drug therapies, such as mefloquine, chloroquine, and so forth (Bloland, 2001). In 2015, the WHO began recommending the use of artemisinin combination therapies for the treatment of malaria (Olumese, 2006). As the name suggests, artemisinin derivatives are the primary component of these therapies and are used in combination with other antimalarial drugs like lumefantrine and mefloquine to slow the progression of multidrug resistance.

Artemisinin is a sesquiterpene lactone peroxide which is primarily extracted from the *Artemisia annua* shrub. Despite its important use in antimalarial therapies, its availability and price fluctuate by as much as 10-fold from season to season owing to fickle agricultural dynamics related to supply and demand (Peplow, 2016). To stabilize the economics of artemisinin, alternate biotechnological routes have long been sought for its production. In 2013, Sanofi launched a semisynthetic artemisinin product after a more than decade-long and $43 M+ initiative to engineer the mevalonate (MVA) pathway in yeast to produce the precursor artemisinic acid

followed by chemical conversion to artemisinin (University of California, Berkeley, 2004). However, at ~$400/kg, the price of semisynthetic artemisinin was still more than double that of its agricultural counterpart, and the process was terminated after just a year of production (Peplow, 2016). As such, there continues to be interest and renewed efforts in developing a more economical biotechnological route for artemisinin production.

In this study, we focus on the production of amorpha-4,11-diene (AMD4,11), the sesquiterpene precursor of artemisinin (Figure S1). AMD4,11 can be oxygenated into artemisinin alcohol, artemisinic aldehyde, and artemisinic acid via a cytochrome P450 enzyme, with the latter two reactions further enhanced by alcohol dehydrogenase and aldehyde dehydrogenase enzymes. While enzymes have been identified for the further biological conversion of artemisinic acid, the semisynthetic process for artemisinin relies on photochemical conversion of artemisinin acid into artemisinin (Kung et al., 2018).

AMD4,11 is produced from farnesyl pyrophosphate (FPP) using amorphadiene synthase (ADS). FPP can be produced from isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP), the building blocks for terpenoids (Kirby & Keasling, 2009) which are generated in nature via two distinct biological pathways. The MVA pathway is native to yeast, and the methylerythritol phosphate (MEP) pathway is native to bacteria, such as *Escherichia coli*. A stoichiometric analysis of these two pathways reveals that the MEP pathway is superior to the MVA pathway for two reasons. First, it has a higher maximum theoretical yield on carbon (1.255 mol glucose per mol of IPP for MEP vs. 1.5 mol glucose per mol of IPP for MVA), making it the most carbon-efficient pathway for terpenoid production (Yadav et al., 2012). Second, the MEP pathway has a zero-net balance of NADH in contrast to the MVA pathway which generates 4 NADH, resulting in a more robust, redox-balanced system (Appendix, Figure S2). It is important to note that the use of bacterial hosts also has clear advantages over eukaryotic platforms like yeast. Because of their faster growth rates, fermentation batch lengths for bacteria are significantly shorter than yeast (2–3 vs. 5–10 days), which enables both higher productivities and improved process economics. As such, there is significant interest in utilizing the MEP pathway with *E. coli* as the host for terpenoid production.

The objective of our study was to identify critical process parameters for optimizing AMD4,11 production via the MEP pathway in *E. coli* in 2 L bioreactors. We studied the impact of monophasic (aqueous) versus biphasic (10% oil) systems on mass transfer and cellular respiration. Through intracellular metabolite analysis and flux balance analysis (FBA), we were able to assess changes in cofactor (NADH/NADPH) and energy (adenosine triphosphate [ATP]) demands which influenced performance and flux through the pathway. Finally, by manipulating relative carbon and oxygen uptake rates, we were able to achieve sustained biomass specific productivities of >1.8 mg/DCW·h and a carbon yield of >5%, thus paving the way for the commercial use of the MEP pathway for terpenoid production.

## 2 | MATERIALS AND METHODS

### 2.1 | Strain *E. coli* AMD

*E. coli* AMD strain was provided by Manus Bio. This strain, derived from *E. coli* MG1655, was engineered for improved flux through the 2-C-methyl-D-erythritol 4-phosphate (MEP) pathway toward the common terpenoid precursors IPP/DMAPP (Kumaran et al., 2019) The ADS gene derived from *A. annua* was integrated to enable the production of AMD4,11.

### 2.2 | Preculture cultivation

Preculture cultivation was started by inoculating 0.8 ml of frozen cell stock (30% glycerol) of *E. coli* strain AMD in 400 ml of batch media. Nonbaffled Erlenmeyer flask with total volume of 2000 ml was used for cultivation. One liter of batch media contained 15 g of glucose, 40 mg of $FeSO_4{\cdot}7H_2O$, 1 ml of 1000× trace elements (10 mg $MnCl_2{\cdot}4H_2O$, 3.2 mg of $ZnSO_4{\cdot}7H_2O$, 1 mg of $CuCl_2{\cdot}2H_2O$, 40 mg of $CaCl_2{\cdot}2H_2O$, 0.50 mg of $H_3BO_3$, and 2 mg of $Na_2MoO_4{\cdot}2H_2O$), 300 mg of $MgSO_4$, 4 mg of thiamine hydrochloride, 3.6 g of $(NH_4)_2SO_4$, 12 g of $KH_2PO_4$, 4.58 g of $K_2HPO_4$, 5 g of yeast extract, and 1 g of citric acid. A total of 0.5 ml of antifoam (Struktol®J 647; Schill+Seilacher) was added to prevent foaming. The shake flask was incubated overnight at 37°C installing 110 RPM and was harvested during exponential growth with optical density (OD) of about 3–4 (detected with 600 nm) before inoculating the 2 L Dasgip bioreactor. Inoculum size was 5% of the bioreactor recipe volume.

### 2.3 | Cultivation in bioreactors

Cultivations in bioreactors were carried out in duplicate installing fed-batch mode. One liter of batch medium for 2 L Dasgip fermenter contained 12 g of glucose, 40 mg of $FeSO_4{\cdot}7H_2O$, 1 ml of 1000× Trace Elements (10 mg $MnCl_2{\cdot}4H_2O$, 3.2 mg of $ZnSO_4{\cdot}7H_2O$, 1 mg of $CuCl_2{\cdot}2H_2O$, 40 mg of $CaCl_2{\cdot}2H_2O$, 0.50 mg of $H_3BO_3$, and 2 mg of $Na_2MoO_4{\cdot}2H_2O$), 300 mg of $MgSO_4$, 4 mg of thiamine hydrochloride, 7.2 g of $(NH_4)_2SO_4$, 12 g of $KH_2PO_4$, 4.58 g of $K_2HPO_4$, 10 g of yeast extract, and 1 g of citric acid. Ten percent (v/v based on initial volume) organic phase of safflower oil was added to the biphasic system. No such addition was made to the control condition yielding a monophasic system. Preliminary tests disclosed that the oil phase does not serve as carbon source for the *E. coli* cells (see Appendix). A total of 0.5 ml of antifoam agent (Struktol®J 647; Schill+Seilacher) was added in the batch media to avoid foaming during the process. When foaming occurred bolus of 0.2 ml of antifoam was added into the fermenter using a syringe at a regular interval. Temperature of 37°C was maintained throughout the process. Before inoculating, pH of the fermenter was adjusted to 7 using 6 N NaOH. pH was maintained at 7 until 42 h using 6 N $NH_4OH$. It also acted as the source of nitrogen along with ammonium sulfate in the batch media. Base was switched to 6 N NaOH after addition of ~180

mmoles of $NH_4OH$ to avoid accumulation of ammonia. Dissolved oxygen ($p_{O2}$) level was maintained at 40% of air saturation for first 24 h and then reduced to 15% linearly over the period of 6 h, which was unchanged for the rest of the process. $p_{O2}$ levels were regulated using agitation cascaded between 300 and 1200 RPM. Aeration was established at 12 L/h throughout the process.

Depletion of glucose in the batch media was indicated by the spike in the $pO_2$ levels. With the exhaustion of glucose, fed-batch process was initiated using a feed solution of 500 g/L of glucose, 31.28 g/L of $KH_2PO_4$, 24.36 g/L of $K_2HPO_4$, 0.48 g/L of $MgSO_4$, 1.6 ml of 1000× trace elements (10 mg $MnCl_2{\cdot}4H_2O$, 3.2 mg of $ZnSO_4{\cdot}7H_2O$, 1 mg of $CuCl_2{\cdot}2H_2O$, 40 mg of $CaCl_2{\cdot}2H_2O$, 0.50 mg of $H_3BO_3$, and 2 mg of $Na_2MoO_4{\cdot}2H_2O$), and 6.4 mg/L of thiamine HCl. As soon as the base switched from 6 N $NH_4OH$ to 6 N NaOH, nitrogen feed consisting of 99 g/L of $(NH)_2SO_4$ and 18.75 ml/L of 25% $NH_4OH$ started to control the concentration of ammonia in the reactor between 30 and 60 mM. Based on glucose concentration in the bioreactor, feed rate was either increased or decreased. Process ran for the estimated fermentation time of 2–3 days. Fermentation profiles are included in the Supporting Information (Figures S7 and S8).

## 2.4 | Glucose and acetate analysis

Samples were withdrawn at a regular interval during the experiment. To separate the supernatant 2 ml of well mixed broth was centrifuged at 20,173*g* at 4°C for 10 min (5430R; Eppendorf). After centrifugation, the aqueous layer of supernatant was carefully pipetted out of the tubes and then filtered through ROTILABO® PVDF, 0.22 µm, 13 mm syringe filters and stored at −70°C for further analysis. A total of 20 µl of the supernatant was added into the reaction tubes of LaboTrace analyzer (Trace Analytics) to measure glucose. Acetate was measured using HPLC Agilent 1200 using a Rezex ROA organic acid column H+ (8%) column (300 × 7.8 mm, 8 µm; Phenomenex). Phosphate interfered with the elution of organic acid during the analysis. Hence, phosphate was subjected to precipitation for all sample and standard before measurement. Thus, 45 µl of 4 M $NH_3$ and 100 µl of 1.2 M $MgSO_4$ were added to 1000 µl sample. After 5 min of incubation, the sample was centrifuged for 5 min at 20,173*g* and room temperature (RT). Then, 500 µl supernatant were transferred to 500 µl of 0.1 M $H_2SO_4$. After thorough mixing and 15 min of incubation at RT, samples were finally centrifuged for 15 min at 20,173*g* at RT. Subsequently, the supernatant loaded for HPLC injection (10 µl injection volume). Separation was performed under isocratic conditions at 50°C (column temperature) for 45 min with 14 mM $H_2SO_4$ as mobile phase at a constant flow rate of 0.6 ml/min. Acetate was detected via an Agilent 1200 Series refractive index detector at 32°C. Six-point calibration curve was used as external reference standard.

## 2.5 | Cell density measurement

Cell pellet obtained from the above probe using centrifugation 20,173*g* (5430R; Eppendorf) at 4°C for 10 min was washed twice with 0.9% NaCl. A total of 0.9% NaCl was added again increasing the total volume to 1 ml which was vortexed to resuspend the pellet. $OD_{600}$ was measured using UV spectrophotometer at 600 nm. For accurate estimation samples were diluted to the optical density between 0.05 and 0.3 before measurement. For cell density measurement, 5 ml of well mixed sample was centrifuged at 1000*g* at 4°C (5430R; Eppendorf) for 10 min and the supernatant was discarded. The pellet was washed twice with 0.9% NaCl and dried at 105°C for 24 h in a convection oven (Heraeus). Pellet was weighed to estimate the cell-dry weight. Based on the OD measurement biomass to OD correlation was calculated to be 0.224.

## 2.6 | Intracellular metabolite measurement

For intracellular measurements around 2 ml sample was withdrawn from the bioreactor using a syringe containing 6 ml of 60% methanol precooled at −30°C for quenching. Syringes were weighed before and after withdrawing the sample. The mixture was transferred to 15 ml falcon tubes and centrifuged at −11°C at 4000*g* (5430R; Eppendorf) for 10 min. Supernatant was discarded and the cell pellet was immediately flash froze in liquid nitrogen. It was then stored in −70°C freezer for further processing.

Based on the biomass concentration in the fermenter, the necessary extraction volume of 50% methanol was calculated to achieve biomass concentration of 50 g/L after extraction. Pellet was resuspended by repetition of 1 min vortexing and 1 min of cooling in the cryostat maintained at −30°C. After the resuspension 100% chloroform equivoluminal to 50% methanol was added. Then, samples were incubated at −20°C for 1 h. This was followed by further incubation at RT for 1 h. It was centrifuged 4000*g* for 10 min at 4°C and the topmost polar layer of aqueous methanol was pipetted out and stored at −70°C freezer for analysis.

ATP quantification was performed on the samples using 1200 Series Agilent HPLC equipped with diode array detector for detection of nucleotides. Method for qualification and quantification on HPLC is described in Löffler et al. (2016). For measurement for MEP pathway intermediates, samples were analyzed on Agilent 1290 Infinity II LC system with 6460 QQQ. Method for qualification and quantification on LC-MS system is described in Zhou et al. (2012).

## 2.7 | AMD4,11 extraction and quantification

For extraction of AMD4,11, 400 µl of sample was transferred into glass tubes. A total of 1600 µl of ethyl acetate (99.9% purity) was added. The mixture was vortexed for 15 min and then centrifuged at 1000*g* for 10 min at RT. The topmost layer of ethyl acetate was pipetted out of the tube and transferred to the glass vials to be stored at −70°C for further analysis.

AMD4,11 quantification of samples was performed on Agilent 5840 GC equipped with flame ionization detector using Restek 5MS-Sil 0.25 mm ID, 0.1 µm df, and 30 m length column for separation. A total of 2 µl of sample was injected into the injection port maintained

at 230°C. Oven temperature was maintained at 90°C for 4 min and then ramped up to 180°C at 5°C/min with the hold time of 3.5 min. It was further increased to 300°C at 50°C/min with the hold time of 5 min. Five-point calibration curve was developed as external reference standard.

## 2.8 | Exhaust gas analysis

Dasgip bioreactors were attached with in line monitoring of exhaust gas composition using a GA4, four channel exhaust gas analyzers equipped with Blue Sens technology. The outlet stream through the fermenter was analyzed for concentration of oxygen, carbon dioxide, and nitrogen in terms of volume percentage.

## 2.9 | Flux balance analysis

We used Cobra Toolbox to perform FBA to optimize ATP4sr reaction, primary ATP synthesis reaction, which was considered as the objective function. iJR904 was used as the base *E. coli* model and further modifications were implemented according to strain construct (Reed et al., 2003). For synthesis of AMD4,11 from farnesyl diphosphate, AMD was added to the model along with export sink Ex_AMD. Reactions associated with quinol oxidase cytochrome *bd* (CYTBD) and pyruvate formate lyase (PFL) respectively were deleted for aerobic condition. Reactions representing quinol oxidase cytochrome $bo_3$ (CYTBO3) and pyruvate dehydrogenase (PDH) were deleted for micro-aerobic condition (Partridge et al., 2007). Isocitrate lyase was deactivated, as glyoxylate shunt cycle is repressed by growth on glucose (Cortay et al., 1989). Reactions associated with idonate oxidation IDOND and IDON2, involving NADH and NADPH respectively, acted as a sink for redox molecules. They were identified as futile reactions and hence were deleted. Our model consisted of 1074 reaction and 762 metabolites. *P/O* ratio of 2 was assumed for phosphorylation. Cellular transport of glucose ($q_{glucose}$), oxygen($q_{O2}$), AMD4,11 ($q_{AMD}$), and specific growth rate ($\mu$) were limited by the constraints quantified from experimental observations. Experimental observations were parsed out from two different timepoint 16.75 and 26 h during the fermentation from both biphasic and monophasic system. Based on the constraints, objective function was optimized to predict fluxes using LP solver.

# 3 | RESULTS

## 3.1 | Impact of overlay addition on fermentation performance

The use of organic overlays during sesquiterpene production is a common practice, as it both facilitates the capture and subsequent purification of volatile products and sequesters product from cells to mitigate potential inhibition or toxicity. However, the addition of oil affects the surface tension of the system, thus modifying air bubble breakage and mass transfer coefficients (Dumont & Delmas, 2003). Because this change directly impacts cellular respiration, we sought to first characterize the effect of oil addition on the fermentation process.

We set up two duplicate conditions in 2L DASGIP bioreactors—a monophasic system with no organic overlay and a biphasic system with the addition of 10% (v/v) safflower oil to facilitate product capture. These bioreactors were ran as fed-batch fermentations with a glucose and phosphate feed which was initiated after the initial depletion of glucose in the batch media at 11 h. A glucose-restricted feeding

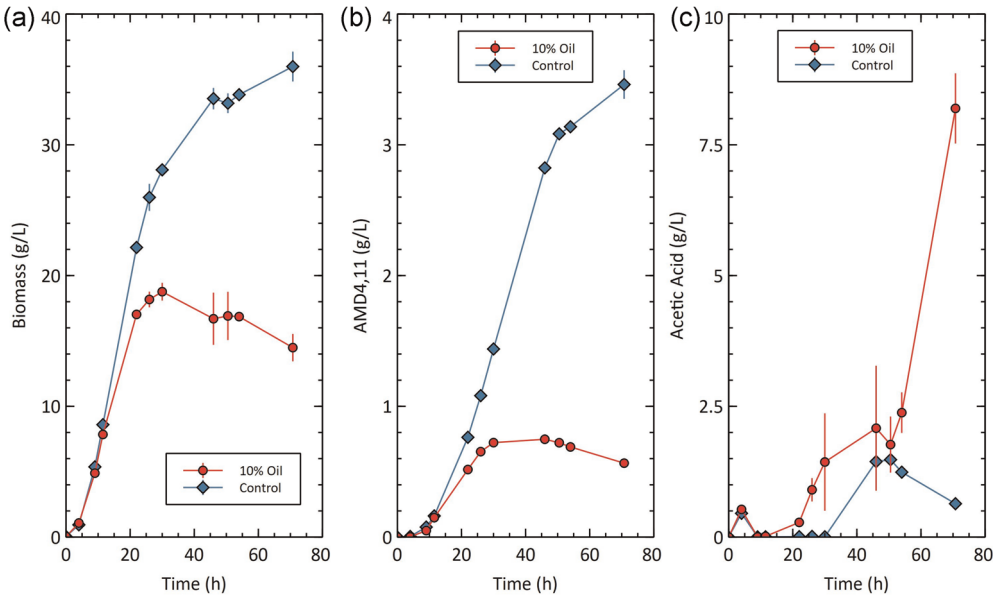



**FIGURE 1** Cell density, production, and acetic acid for series 1. (a) Biomass concentration (g/L), (b) AMD4,11 titer (g/L), and (c) acetic acid concentration (g/L) for biphasic (10% oil) and monophasic (control) in series 1. AMD4,11, amorpha-4,11-diene [Color figure can be viewed at wileyonlinelibrary.com]

strategy was implemented to avoid glucose accumulation and by-product formation (Han et al., 1992), and all fermenters were subjected to the same feed rates.

As seen in Figure 1, we observed differences in both biomass and AMD4,11 accumulation between the two conditions, with the monophasic system outperforming the biphasic system with respect to both overall titer and cell specific productivities. Peak biomass concentrations were 40% higher in the monophasic system compared to the biphasic system (26 vs. 18 g DCW/L), and AMD4,11 titers were almost sixfold higher in the monophasic system at 71 h (3.5 vs. 0.6 g/L; Figure 1a,b).

Although we observed large differences in overall performance, our time course analysis reveals that the deviations between the two systems began with the onset of the fed-batch phase. Despite having established the same feed rates for both systems, biomass specific uptake rates of glucose were higher in the biphasic approach than in the monophasic system (0.29 vs. 0.23 g/DCW·h at 17 h; Figure 2a). Specific oxygen uptake rates were also 20–25% higher for the biphasic system throughout the fed-batch portion of the run (Figure 2b). Interestingly, these higher glucose and oxygen uptake rates correlated with both lower biomass formation and a sharp decline in the specific productivity of AMD4,11. For the oil-free system, AMD4,11 productivities remained at around 4 mg/DCW·h in growing cells and then continued at a low level of 1–2 g mg/DCW·h until the end of the fermentation. In contrast, we observed a large drop from 5.5 mg/DCW·h after only 19 h into the run for the biphasic condition (Figure 2c).

## 3.2 | Intracellular ATP and metabolite measurements

To gain insight into the dynamics of the system, we measured the intracellular levels of three key metabolites—ATP, 1-deoxy-D-xylulose (DOX), and (*E*)-4-hydroxy-3-methyl-but-2-enyl pyrophosphate (MEcPP), the latter two of which serve as important indicators of MEP pathway flux and functionality. Although we attempted to collect data from four time points throughout the run, the final time point (71 h) was excluded for the biphasic system due to severe foaming which impaired our ability to obtain representative samples.

As seen in Figure 3a, we again observed significant deviations between the two conditions tested, with biomass specific ATP levels between 2- and 12-fold higher for the biphasic system compared to the monophasic system throughout the run (Figure 3a). The dynamics of the MEP pathway also varied based on both DOX and MEcPP measurements. DOX is the dephosphorylated product of 1-deoxy-D-xylulose 5-phosphate (DOXP), the first intermediate of

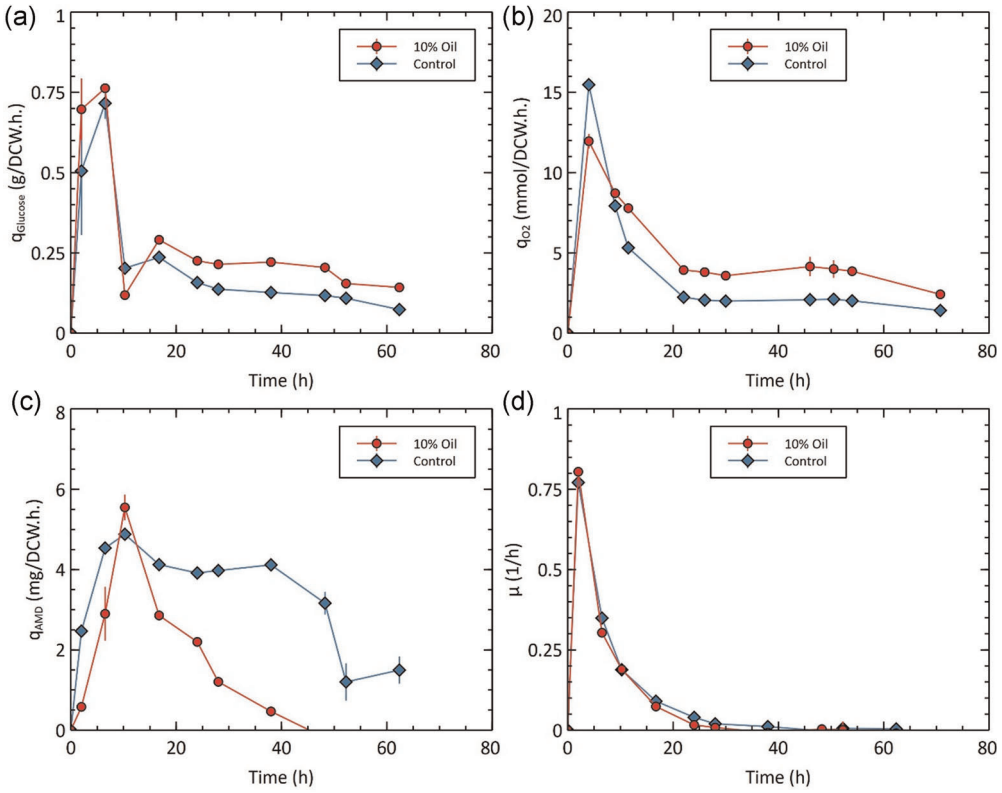



**FIGURE 2** Biomass specific dynamics for series 1. Biomass specific uptake rate of (a) glucose ($q_{glucose}$) and (b) oxygen ($q_{O2}$). (c) Specific productivity of amorpha-4,11-diene ($q_{AMD}$) and (d) specific growth rate ($\mu$) for biphasic (10% oil) and monophasic (control) in series 1. Units: $q_{glucose}$, g/DCW·h; $q_{O2}$, mmol/DCW·h; $q_{AMD}$, mg/DCW·h; $\mu$, 1/h [Color figure can be viewed at wileyonlinelibrary.com]

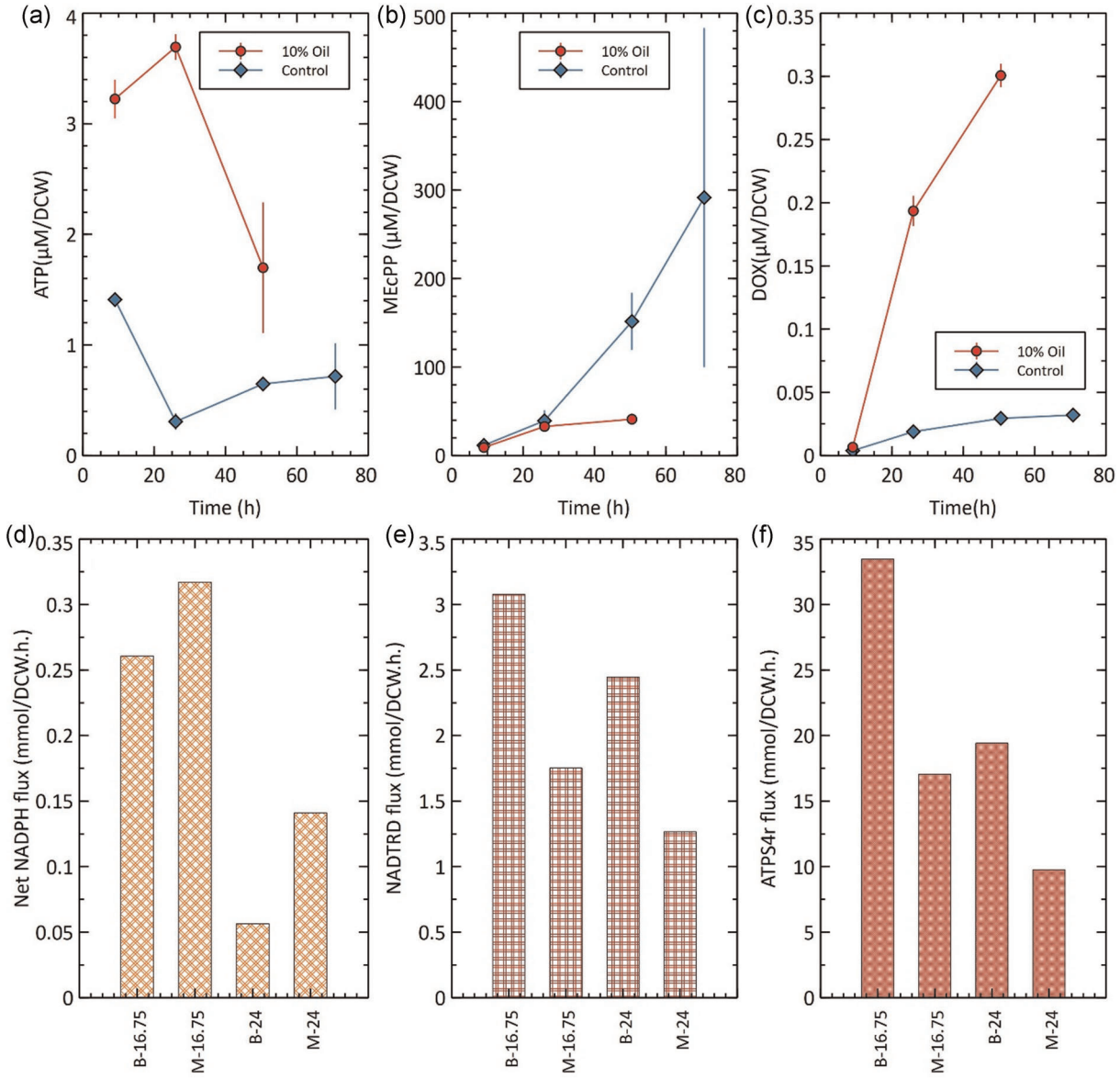



**FIGURE 3** Intracellular measurements for series 1 and energy management and redox balance for FBA. Intracellular concentrations of (a) ATP, (b) MEcPP and (c) DOX for biphasic (10% oil) and monophasic (control) in the first row of diagram; unit: μM/DCW. Second row of diagram represents FBA results for (d) net nicotinamide adenine dinucleotide phosphate (NADPH) flux illustrating the availability of NADPH. (e) Fluxes through NADTRD converting NADPH to NADH, $NADPH + NAD^{+} = NADP^{++}$ NADH. (f) ATP synthesis flux is indicated by ATP4sr reaction. B-16.75 and B-24 represent results for 16.75 and 24 h for biphasic (10% oil) system. M-16.75 and M-24 represent results for 16.75 and 24 h for monophasic (control) system; Unit: mmol/DCW·h. ATP, adenosine triphosphate; DOX, 1-deoxy-D-xylulose; FBA, flux balance analysis; MEcPP, 2-C-methyl-D-erythritol 2,4-cyclodiphosphate; NADTRD, NAD transhydrogenase [Color figure can be viewed at wileyonlinelibrary.com]

the MEP pathway, and its level gives clues into both the amount of carbon flowing into the MEP pathway and the ability of subsequent enzymes in the pathway to pull this carbon downstream (Figure 4). MEcPP is a product of IspF and is a double phosphorylated intermediate. Its accumulation also serves as a proxy for further bottlenecks downstream in the pathway. In Figure 3c, we see that DOX is at least 10-fold higher for the biphasic system compared to the monophasic system at both 25 and 50 h. In contrast, MEcPP deviates only after 26 h, with the monophasic system showing a sharp increase to 151.5 μM/DCW at 50 h, and the biphasic system showing a steady level of ~40 μM/DCW throughout (Figure 3b). We note that the large error bars seen at 71 h is reflective of foaming that was observed at this stage of the fermentation.

## 3.3 | Stoichiometric modeling to estimate carbon distribution and cofactor demands

To estimate the distribution of carbon flux in our two systems, FBA was performed for two time points (16.75 and 24 h) using

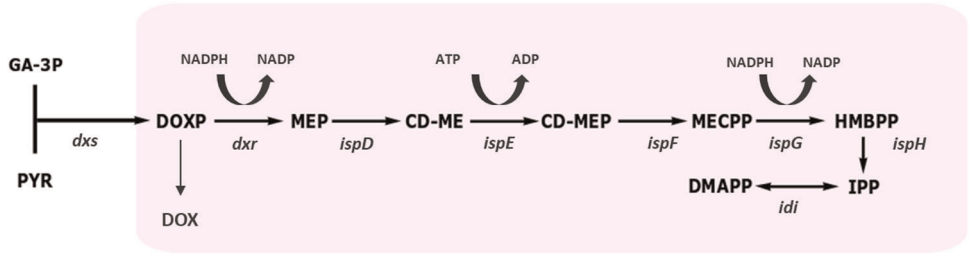



**FIGURE 4** MEP pathway. GA-3P reacts with pyruvate to from DOXP. If accumulating, DOXP gets converted into DOX. Metabolites: MEP, CD-ME, CD-MEP, MECPP, HMBPP, IPP, and DMAPP. Enzymes: *dxs*, *dxr*, *ispD*, *ispE*, *ispF*, *ispG*, *ispH*, and *idi*. CD-ME, 4-diphosphocytidyl-2-*C*-methylerythritol; CD-MEP, 4-diphosphocytidyl-2-*C*-methyl-D-erythritol 2-phosphate; DMAPP, dimethylallyl pyrophosphate; DOX, 1-deoxy-D-xylulose; DOXP, 1-deoxy-D-xylulose 5-phopshtase; *dxr*, DXP reductoisomerase; *dxs*, DOXP synthase; GA-3P, glyceraldehyde 3-phosphate; HMBPP, (*E*)-4-Hydroxy-3-methyl-but-2-enyl pyrophosphate; *idi*, isopentenyl pyrophosphate isomerase; IPP, isopentenyl pyrophosphate; *ispD*, 2-*C*-methyl-D-erythritol-4-phosphate cytidylyltransferase; *ispE*, 4-diphosphocytidyl-2-*C*-methyl-D-erythritol kinase; ispF, 2-*C*-methyl-D-erythritol 2,4-cyclodiphosphate synthase; *ispG*, HMB-PP synthase; *ispH*, HMB-PP reductase; MECPP, 2-*C*-methyl-D-erythritol 2,4-cyclodiphosphate; MEP, 2-*C*-methylerythritol 4-phosphate [Color figure can be viewed at wileyonlinelibrary.com]

experimentally-generated data ($\mu$, $q_{O2}$, $q_{glucose}$, $q_{AMD}$) as model inputs/constraints (Table 1). Since the MEP pathway consumes 1 ATP but does not provide opportunities for regeneration, we chose to maximize the ATP synthesis reaction ATPS4r as our objective function, resulting in the model shown in Figure 5. For comparison, fluxes including $q_{O2}$ were normalized with respect to glucose uptake rates $q_{glucose}$.

Although fluxes in glycolysis were similar in all conditions irrespective of whether oil was present or not, we observed more stark differences in TCA and MEP pathway activity (Figure 5). During the growth phase (16.75 h), the biphasic model was found to have elevated TCA activity and lowered MEP fluxes compared to the monophasic system. However, at 24 h, TCA activity dropped by 96% for the biphasic system, coinciding with the installation of oxygen-limiting conditions. It is well known that the aerobically active PDH enzyme has reduced activity in favor of PFL under microaerobic and anaerobic conditions (Partridge et al., 2007). Thus, the conversion of pyruvate to acetyl-CoA via PFL is likely to occur at 16.75 h and at 24 h in the biphasic system when $pO_2$ was ~0% of air saturation. At 24 h, higher TCA fluxes were generally correlated with higher MEP fluxes and AMD4,11 productivities.

## 3.4 | Calculating net NADPH flux

By summing up all reactions involved in the production and consumption of NADPH, we are able to derive a net NADPH flux for our monophasic and biphasic systems. Interestingly, the net NADPH formation of growing cells was 17.6% lower with oil than without oil, a trend that was further amplified at 24 h (Figure 3d). Inversely, fluxes via NAD transhydrogenase (Figure 3e) were highest in the biphasic system at both time points (3.08 mmol/DCW·h at 16 h and 2.5 mmol/DCW·h at 24 h). Fluxes for the ATP synthesis reaction, ATPS4r (Figure 3f) were highest during the growth phase (16 h) in the biphasic system (33.47 mmol/DCW·h) but were reduced by almost half for the other three data points.

## 3.5 | Optimizing glucose and oxygen uptake rates to improve performance of biphasic system

Based on our previous FBA analysis, we postulated that a high $q_{O2}/q_{glucose}$ ratio may lead to limited availability of both pyruvate and NADPH, two critical substrates for high MEP pathway functionality. We therefore anticipated that balancing the biomass specific glucose and oxygen uptake rates may improve cell productivities in biphasic systems. To test this, we performed another series of monophasic and biphasic fed batch fermentations. Although the monophasic fermentations were conducted as before, glucose uptake rates were lowered for the biphasic system by feeding glucose at a 10% reduced rate. The aeration regime was left unchanged.

As before, glucose in the batch media was depleted after about 11 h for both systems, after which glucose and phosphate feeding

**TABLE 1** Parameter definition for flux balance analysis constraints applied for flux balance analysis

| Condition | System | EFT (h) | $q_{O2}$ (mmol/DCW·h) | $q_{glucose}$ (mmol/DCW·h) | $q_{AMD}$ (mmol/DCW·h) | $\mu$ (1/h) |
|---|---|---|---|---|---|---|
| B-16.75 | Biphasic | 16.75 | 5.38 | 1.61 | 0.014 | 0.074 |
| M-16.75 | Monophasic | 16.75 | 3.42 | 1.31 | 0.02 | 0.09 |
| B-24 | Biphasic | 24 | 3.52 | 1.25 | 0.011 | 0.016 |
| M-24 | Monophasic | 24 | 2.09 | 0.87 | 0.019 | 0.039 |

*Note*: Values are obtained from the experimental observation for both monophasic and biphasic systems.

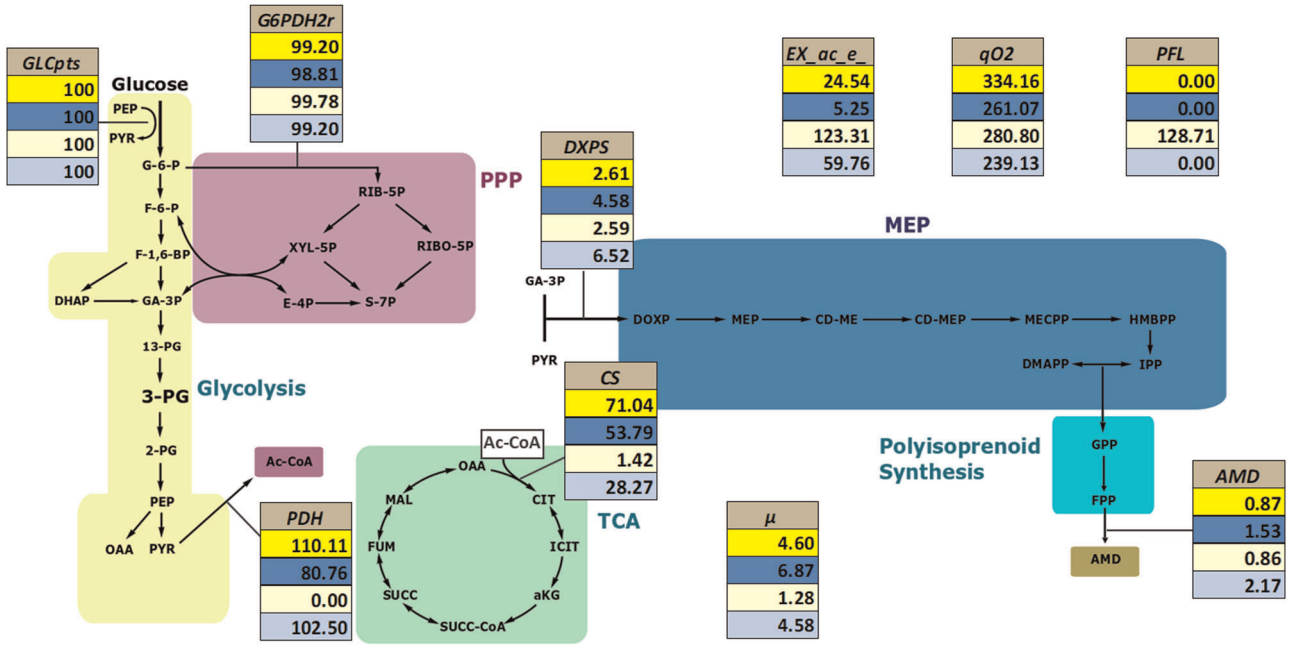



**FIGURE 5** Flux balance analysis results for series 1. Flux prediction based on the constraints applied using results of experiment 1. Flux through individual reactions are depicted for B-16.75, M-16.75, B-24, and M-24 in blue, yellow, sky blue, and lemon-yellow cells, respectively. Reactions: GLCpts, G6PDH2r, PDH, CS, DXPS, AMD, EX_ac_e_, $q_{O2}$, PFL, and $\mu$. All fluxes are in mmol/DCW·h. Fluxes are normalized by specific glucose uptake rate ($q_{glucose}$). $\mu$, specific growth rate; AMD, amorpha-4,11-diene synthase; CS, citrate synthase; DXPS, DOXP synthase; EX_ac_e_, acetate export; G6PDH2r, glucose-6-phosphate dehydrogenase; GLCpts, glucose transport via PTS system; PDH, pyruvate dehydrogenase; PFL, pyruvate formate lyase; $q_{O2}$, biomass specific oxygen uptake rate [Color figure can be viewed at wileyonlinelibrary.com]

was initiated. This time, we observed similar performance between both conditions, with the key parameters, including biomass production, glucose uptake rates, and oxygen uptake rates tracking closely throughout the run (Figure 6). The biphasic fermentations reached 3.5 g/L AMD in 64 h with a sustained cell specific productivity of ~4 mg/DCW·h throughout most of the run, rivaling the performance of the monophasic system (Figure 6c). Similar trends were also seen for intracellular ATP, DOX, and MEcPP measurements as depicted in Figure 7, with DOX and MEcPP levels rising steadily over the course of the fermentation.

In light of above hypothesis, we conducted series 3 and 4 to further consolidate our findings. Series 3 was conducted with mono- and biphasic (10% oil) compositions in duplicates. Both settings were investigated to mirror specific glucose uptake profiles of the biphasic system of series 1 (~0.24 g/DCW/h). As a result, both settings operated at $q_{O2}$ around 5 mmol/DCW·h for most part of the cultivation (Figure S5). As expected, amorphadiene titer was impacted negatively with the monophasic system achieving $1.09 \pm 0.02$ g/L and biphasic system producing only $0.74 \pm 0.06$ g/L (Figure S4).

For complementation, series 4 was conducted checking the biphasic approach by installing 2.5% and 10% oil in duplicates. We aimed at mimicking specific glucose uptake rates with those of series 2. Highest $q_{O2}$ for all the three systems was around 2 mmol/DCW·h after 20 h. At the end, that is, after 68.5 h, the monophasic approach reached $3.37 \pm 0.16$ g/L, 10% oil was $3.82 \pm 0.35$ g/L and 2.5% oil system was $3.38 \pm 0.34$ g/L. Figure 8 gives an outline of all the four studies illustrating a relationship between $q_{glucose}$, $q_{O2}$ attained after 20 h of elapsed fermentation time and final AMD4,11 titer of the cultivation. Two clusters were formed indicating the correlation between $q_{glucose}$ and $q_{O2}$. $q_{O2}$ of less than ~2.5 mmol/DCW·h was identified as a region with balanced NADPH supply, thus improving its availability for MEP pathway (indicated by checkered area). On the contrary, $q_{O2}$ above 3 mmol/DCW·h drained NADPH into NADH lowering its accessibility for AMD4,11 production. Lowering the specific glucose update rate resulted in lower $q_{O2}$ and improvement in AMD4,11 titer.

## 4 | DISCUSSION

### 4.1 | Use of an organic overlay leads to glucose overfeeding

Because of the hydrophobicity of sesquiterpene molecules, an organic oil overlay is commonly added to the fermentation broth to capture the products in situ from the aqueous phase. To maintain the same overall reactor volume, we reduced the aqueous volume of the biphasic system by 10%, which resulted in a proportional increase in the concentration of glucose provided in either the batch media or by the feed. Our first set of fermentations revealed no differences in growth, AMD4,11 production, or glucose uptake rates during the initial batch phase (Figure 1a,b), presumably because rates of

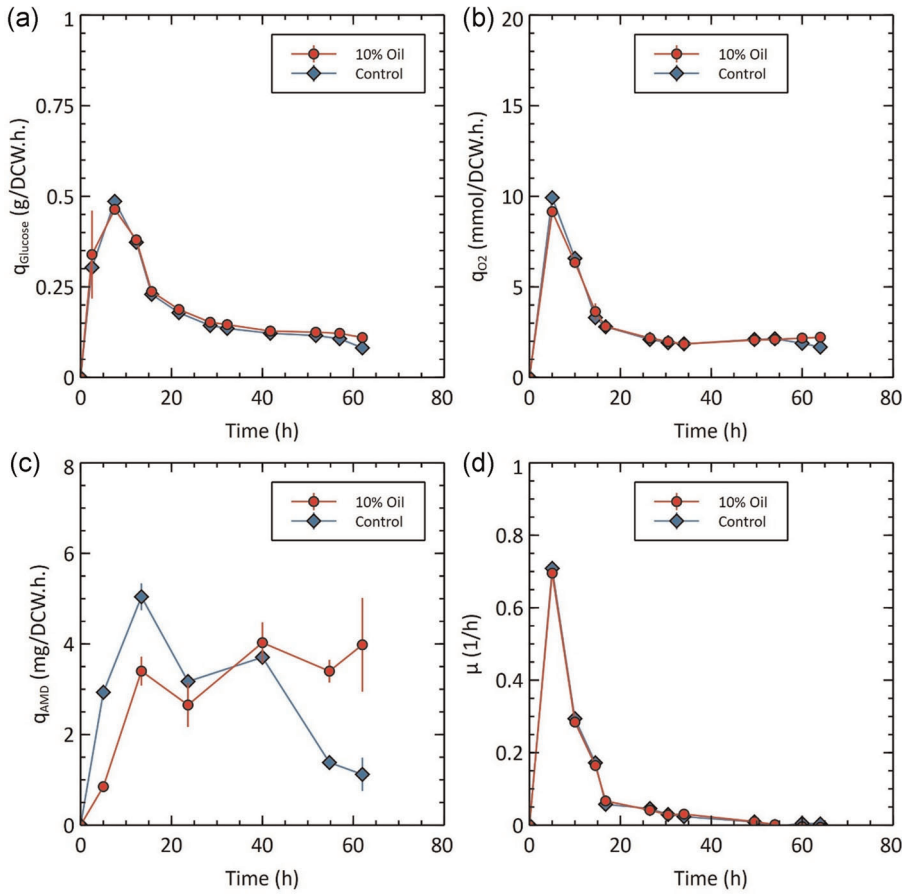



**FIGURE 6** Biomass specific dynamics for series 2. Biomass specific uptake rate of (a) glucose ($q_{glucose}$) and (b) oxygen ($q_{O2}$). (c) Specific productivity of amorpha-4,11-diene ($q_{AMD}$) and (d) growth rate ($\mu$) for biphasic (10% oil) and monophasic (control) in series 2. Units: $q_{glucose}$, g/DCW·h; $q_{O2}$, mmol/DCW·h; $q_{AMD}$, mg/DCW·h; $\mu$, 1/h [Color figure can be viewed at wileyonlinelibrary.com]

nutrient consumption were more impacted by the utilization of yeast extract, as cells prefer to import amino acids rather than synthesizing it (Lyubetskaya et al., 2006). However, we observed a shift upon the start of glucose feeding, with the biphasic fermentations displaying 20% higher specific glucose uptake rates than the oil-free reactors despite maintaining the same volumetric glucose feed rates for both systems.

## 4.2 | Glucose overfeeding increases oxygen demands

The rise in specific glucose uptake rates within the biphasic system coincided with increased specific oxygen demands, reduced growth rates, and declining AMD4,11 productivities (Figure 2). Interestingly, the rise in oxygen uptake rates did not just increase proportionally with the glucose uptake rates but resulted in a higher $q_{O2}/q_{Glucose}$ ratio for the biphasic system. As predicted by our FBA model, this increase leads to a fundamental shift in the distribution of carbon, resulting in lower flux at the pyruvate node which is critical for high MEP pathway functionality (Figure 5). During aerobic respiration, more carbon enters the TCA cycle, pulling from the essential substrate pyruvate and leading to the generation of $CO_2$ which decreases the carbon yield for both biomass and AMD4,11 production. When the biphasic system shifts into a microaerobic state at 24 h (SSB-2), the flux into the TCA cycle is lowered by 96%. Pyruvate is partially diverted toward acetate production, resulting in a fivefold increase in acetate concentration and a 12-fold higher accumulation for the biphasic system compared to the oil-free condition (Figure 1c). These findings reflect the well-known property of *E. coli* to produce acetate once cells surpass threshold specific rates of glucose consumption (Eiteman & Altman, 2006). The presence of acetate coupled with oxygen limitation adversely affects the specific growth rate of *E. coli* (Luli & Strohl, 1990).

## 4.3 | Increased oxygen uptake also lowers NADPH availability

Under aerobic conditions, oxygen is the preferred electron acceptor, linking electron uptake from NADH with the stoichiometric formation of $NAD^+$, ATP, and $H_2O$. Because of the high reactivity of oxygen, its utilization should be safely controlled to prevent the formation of nonwanted radicals and toxic compounds (Imlay, 2013). Accordingly, increased oxygen uptake rates require a proper supply of NADH which is predominately produced in the TCA cycle. In cases of high NADH demand and insufficient supply, cells are also known to use their native transhydrogenase activity to convert NADPH to NADH (Spaans et al., 2015). In Figure 3d, we show that the net NADPH flux availability was 18% lower for the biphasic system

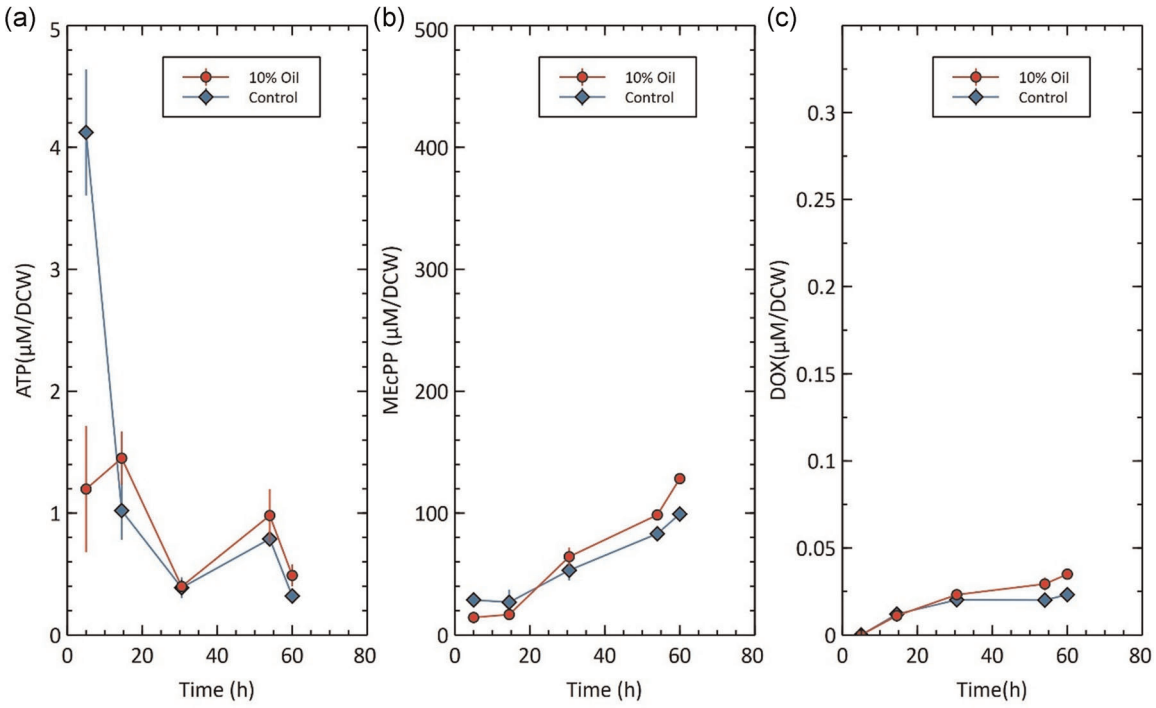



**FIGURE 7** Intracellular measurements for series 2. Intracellular concentrations of (a) ATP, (b) MEcPP, and (c) DOX for biphasic (10% oil) and monophasic (control). ATP, adenosine triphosphate; DOX, 1-deoxy-D-xylulose; MEcPP, 2-C-methyl-D-erythritol 2,4-cyclodiphosphate [Color figure can be viewed at wileyonlinelibrary.com]

compared to the oil-free approach. As such, AMD4,11 production accounted for a higher proportion of the NADPH flux in the biphasic system (32% vs. 19%). This suggests that increased glucose feeding and uptake rates in the biphasic system also had the effect of decreasing the availability of NADPH required for AMD4,11 production.

This hypothesis is further supported by the pattern of accumulation observed for key MEP pathway intermediates. As seen in Figure 3, the biphasic fermentations showed significant accumulation of DOX, the dephosphorylated derivative of DOXP. The subsequent conversion of DOXP to MEP via the Dxr enzyme in the MEP pathway requires NADPH as a cofactor; thus, the 10-fold higher accumulation of DOX in our biphasic system suggests limited NADPH availability in this system. As seen in both Figures 2c and 5, this shortage ultimately results in much lower AMD4,11 titers and productivities.

## 4.4 | Reduction in NADPH availability coincides with increased ATP supply

As seen in Figure 3e, we predict that the reduction in NADPH availability is caused by elevated NADH transhydrogenase activity, which consumes between 69% and 82% of NADPH in the biphasic system compared to 51–62% in the oil-free approach. The resulting NADH enters respiration, which leads to both higher ATP formation rates (Figure 3f) and elevated levels of ATP (Figure 3a) in the

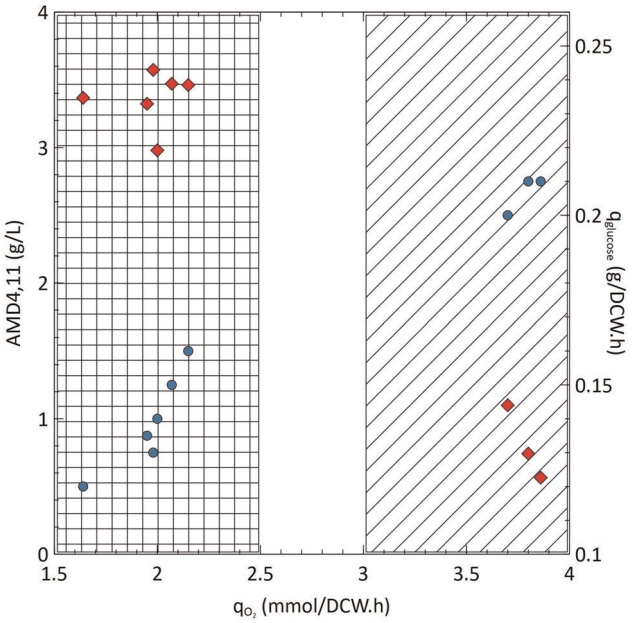



**FIGURE 8** AMD4,11 titer and fed-batch substrate uptake kinetics for series 1, 2, 3, and 4. Final AMD4,11 titer in g/L (◊) and biomass specific glucose uptake rates, $q_{glucose}$ in g/DCW·h (○) for corresponding specific oxygen uptake rate, $q_{O2}$ in mmol/DCW·h installed after 20 h in series 1, 2, 3, and 4. Checkered area ($q_{O2} < 2.5$) corresponds to balanced NADPH supply, while striped area ($q_{O2} > 3$) corresponds to NADPH drainage to NADH. AMD4,11, amorpha-4,11-diene [Color figure can be viewed at wileyonlinelibrary.com]

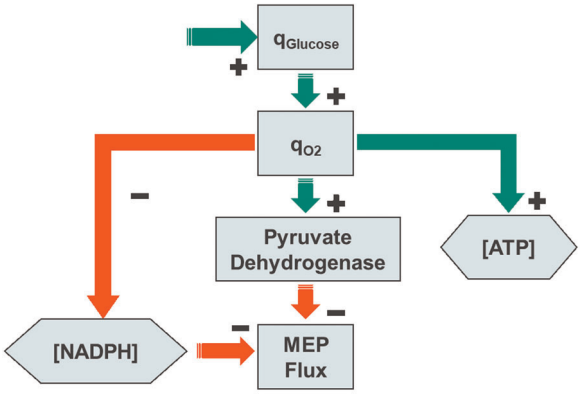



**FIGURE 9** Schematic illustration of dependence of flux dynamics on specific uptake rate of glucose ($q_{glucose}$). Green arrows indicate positive effect and red indicate negative effect. Hexagonal boxes represent concentrations while rectangles represent rates [Color figure can be viewed at wileyonlinelibrary.com]

biphasic system. The decrease in ATP concentrations toward 50 h mirrors the decrease in $q_{O2}$, resulting in inefficient ATP synthesis due to the microaerobic condition (Chapman et al., 1971).

## 4.5 | Summary of key process parameters influencing AMD4,11 production

Through these fermentations, we have established that increase in $q_{O2}/q_{glucose}$ negatively impacts MEP pathway flux and hence, AMD4,11 productivities. Taking this data and observations in aggregate, we propose a series of biochemical mechanisms that explains this phenomenon as shown in Figure 9.

We observed that the presence of an organic overlay reduced the aqueous volume in the bioreactor, thus leading to a slight increase in glucose concentrations available to the cells during the fed-batch phase despite the installation of equivalent feed rates. This led to an increase in glucose uptake rates and a corresponding rise in oxygen uptake rates, which positively impacted flux through PDH into the TCA cycle. This shift in carbon flux had the effect of reducing the amount of pyruvate for the MEP pathway. Simultaneously, higher oxygen uptake rates led to a higher utilization of NADH which was partially provided by the transhydrogenase-mediated conversion of NADPH to NADH. Lower overall availability of NADPH also had the effect of reducing the flux through the MEP pathway and into AMD4,11.

The observed phenomenon is specific for the MEP pathway, as the MVA pathway differs in its use of substrates and cofactors and results in the generation of excess NADH (Yadav et al., 2012). Interestingly, other groups have observed synergies between the MEP and MVA pathways and have shown the ability to increase flux through the MEP pathway in NADPH-limiting conditions when both are expressed together in *E. coli* (Yang et al., 2016). This finding is in agreement with our proposed mechanism in Figure 9.

## 4.6 | Maximizing yield and productivity by balancing glucose and uptake rates

In a subsequent set of fermentations, we found that lowering the glucose feed rates in our biphasic system by 10% (proportional to the increase seen in oil addition) was successful in completely recovering AMD4,11 production to levels that matched or even exceeded the monophasic condition (Figure 10b). All other measured and

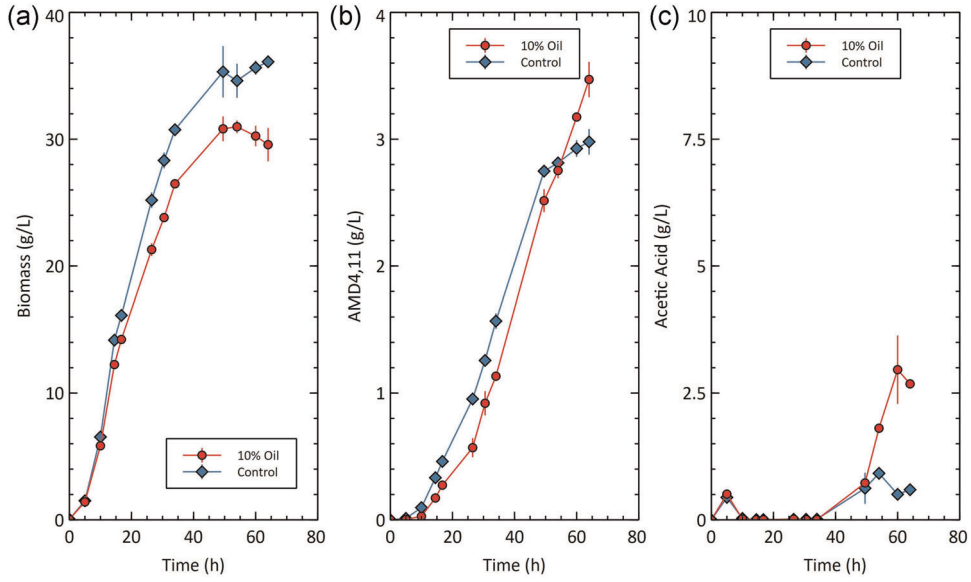



**FIGURE 10** Cell density, production, and acetic acid for series 2. (a) Biomass concentration (g/L), (b) AMD4,11 titer (g/L), and (c) acetic acid concentration (g/L) for biphasic (10% oil) and monophasic (control) in series 2. AMD4,11, amorpha-4,11-diene [Color figure can be viewed at wileyonlinelibrary.com]

calculated parameters also followed similar trends between both monophasic and biphasic conditions (Figures 10 and 6). Importantly, we observed a six-fold increase in AMD4,11 titers, an almost seven-fold increase in conversion yield, and a 67% drop in acetate accumulation in our biphasic fermentations from the first to second run, showing that we were able to relieve the constraints on both pyruvate and NADPH availability to enable higher fluxes through the MEP pathway to AMD4,11 (Figure 10c). Since we still see a rise in MEcPP throughout the course of the fermentation (Figure 7b), we believe that additional strain engineering at either the IspG node or in downstream reactions may be needed to completely pull flux through the MEP pathway (as a side note, although we observed a slightly reduced AMD4,11 titer for our monophasic control during the second run, this was attributed to a short, unintended limitation in oxygen which occurred at 48 h).

Our results compare favorably to other examples of AMD4,11 production in the literature. Tsuruta et al. (2009) used the nonnative MVA pathway in *E. coli* and achieved 1.94 mg/DCW·h of AMD4,11 in a process lasting 160 h. Our approach reached 2.44 mg/DCW·h of AMD4,11 in just 60 h. Westfall et al. (2012) were able to attain production of 5.5 g/L of AMD4,11 in 95 h with a 3.23 C-mol% yield on glucose in *Saccharomyces cerevisiae*. If we extrapolate the volumetric productivity of our approach out to 95 h, we estimate reaching titers of about 5.2 g/L AMD4,11. By installing the proper $q_{O2}/q_{glucose}$ conditions in our fermentations, our *E. coli* AMD4,11 producer achieved a 5.08% C-mol yield which outperforms the results using the MVA pathway.

## 5 | CONCLUSION

By performing an in-depth multilayered characterization of our work that merged standard fermentation monitoring with metabolomics and stoichiometric modeling (FBA), we were able to unravel the complex metabolic interplay of glucose feeding, oxygen uptake, carbon redistribution, and cofactor availability on MEP pathway functionality and sesquiterpene production. Although the improvements enabled by process parameter tuning were demonstrated for AMD4,11 production, these findings are likely to apply generally to the production of other terpenoid products because they address fundamental problems related to the regulation and flux through the MEP pathway. We anticipate that further tuning of these key parameters will allows us to the tap the full potential of the most carbon-efficient pathway for terpenoid production.

### ACKNOWLEDGMENTS

The authors gratefully acknowledge Manus Bio Inc. (Cambridge, USA) for funding this study. We thank Maria Aniolek for her assistance with HPLC, as well as Alexander Dietrich, Andreas Freund and Salaheddine Laghrami for assistance with lab operations. Open Access funding enabled and organized by Projekt DEAL.

### CONFLICT OF INTERESTS

The authors have financial interests in Manus Bio, Inc.

### AUTHOR CONTRIBUTIONS

Vikas Patil and Ralf Takors conceived and designed the study. Vikas Patil conducted the study, analyzed the datasets, and drafted the manuscript. Ralf Takors and Christine N. S. Santos corrected the manuscript. All authors approved the final manuscript.

### NOMENCLATURE

| | |
|---|---|
| $q_{glucose}$ | biomass specific glucose uptake rate (g/DCW·h) |
| $q_{O2}$ | biomass specific glucose uptake rate (mmol/DCW·h) |
| $q_{AMD}$ | biomass specific productivity of amorpha-4,11-diene (mg/DCW·h) |
| $\mu$ | specific growth rate (1/h) |
| $p_{O2}$ | dissolved oxygen (% saturation) |

### ABBREVIATIONS

| | |
|---|---|
| AMD4,11 | amorpha-4,11-diene |
| ATP | adenosine triphosphate |
| DMAPP | dimethylallyl pyrophosphate |
| DOX | 1-deoxy-D-xylulose |
| DOXP | 1-deoxy-D-xylulose 5-phopshtase |
| IPP | isopentenyl pyrophosphate |
| MECPP | 2-C-methyl-D-erythritol 2,4-cyclodiphosphate |
| NADH | nicotinamide adenine dinucleotide reduced |
| NADPH | nicotinamide adenine dinucleotide phosphate reduced |
| $NADP^+$ | nicotinamide adenine dinucleotide phosphate oxidized |
| $NAD^+$ | nicotinamide adenine dinucleotide oxidized |

### ORCID

*Ralf Takors* http://orcid.org/0000-0001-5837-6906

### REFERENCES

Bloland, P. (2001). *Drug resistance in malaria*. Geneva: World Health Organization.

Chapman, A. G., Fall, L., & Atkinson, D. E. (1971). Adenylate energy charge in *Escherichia coli* during growth and starvation. *Journal of Bacteriology*, *108*(3), 1072–1086. https://doi.org/10.1093/jmammal/gyv190

Cortay, J. C., Bleicher, F., Duclos, B., Cenatiempo, Y., Gautier, C., Prato, J. L., & Cozzone, A. J. (1989). Utilization of acetate in *Escherichia coli*: Structural organization and differential expression of the ace operon. *Biochimie*, *71*(9-10), 1043–1049.

Dumont, E., & Delmas, H. (2003). Mass transfer enhancement of gas absorption in oil-in-water systems: A review. *Chemical Engineering and Processing: Process Intensification*, *42*(6), 419–438. https://doi.org/10.1016/S0255-2701(02)00067-3

Eiteman, M., & Altman, E. (2006). Overcoming acetate in *Escherichia coli* recombinant protein fermentation. *Trends in Biotechnology*, *24*, 530–536. https://doi.org/10.1016/j.tibtech.2006.09.001.

Han, K., Lim, H. C., & Hong, J. (1992). Acetic acid formation in *Escherichia coli* fermentation. *Biotechnology and Bioengineering*, *39*(6), 663–671. https://doi.org/10.1002/bit.260390611

Imlay, J. (2013). The molecular mechanisms and physiological consequences of oxidative stress: Lessons from a model bacterium. *Nature Reviews Microbiology*, *11*, 443–454. https://doi.org/10.1038/nrmicro3032

Kirby, J., & Keasling, J. D. (2009). Biosynthesis of plant isoprenoids: Perspectives for microbial engineering. *Annual Review of Plant Biology*, *60*(1), 335–355. https://doi.org/10.1146/annurev.arplant.043008.091955

Kumaran, A. P., Lim, R., Donald, J., Tseng, H., Santos, C., & Philippe, R. Metabolic engineering for microbial production of terpenoid products. US Patent No. 10480015 B2, November 19, 2019.

Kung, S. H., Lund, S., Murarka, A., McPhee, D., & Paddon, C. J. (2018). Approaches and recent developments for the commercial production of semi-synthetic artemisinin. *Frontiers in Plant Science*, *9*, 1–7. https://doi.org/10.3389/fpls.2018.00087

Löffler, M., Simen, J. D., Jäger, G., Schäferhoff, K., Freund, A., & Takors, R. (2016). Engineering *E. coli* for large-scale production—Strategies considering ATP expenses and transcriptional responses. *Metabolic Engineering*, *38*, 73–85. https://doi.org/10.1016/j.ymben.2016.06.008

Luli, G. W., & Strohl, W. R. (1990). Comparison of growth, acetate production, and acetate inhibition of *Escherichia coli* strains in batch and fed-batch fermentations. *Applied and Environmental Microbiology*, *56*(4), 1004–1011. https://doi.org/10.1128/aem.56.4.1004-1011.1990)

Lyubetskaya, A. V., Rubanov, L. I., & Gelfand, M. S. (2006). Use of the flux model of amino acid metabolism of *Escherichia coli*. *Biochemistry (moscow)*, *71*(11), 1256–1260.

Olumese, P. (2006). *Guidelines for the treatment of malaria*. Geneva: World Health Organization.

Partridge, J. D., Sanguinetti, G., Dibden, D. P., Roberts, R. E., Poole, R. K., & Green, J. (2007). Transition of *Escherichia coli* from aerobic to micro-aerobic conditions involves fast and slow reacting regulatory components. *Journal of Biological Chemistry*, *282*(15), 11230–11237. https://doi.org/10.1074/jbc.M700728200

Peplow, M. (2016). Synthetic biology's first malaria drug meets market resistance. *Nature News*, *530*, 389–390. https://www.nature.com/news/synthetic-biology-s-first-malaria-drug-meets-market-resistance-1.19426

Reed, J. L., Vo, T. D., Schilling, C. H., & Palsson, B. O. (2003). An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). *Genome Biology*, *4*(9), R54.

Spaans, S. K., Weusthuis, R. A., van der Oost, J., & Kengen, S. W. M. (2015). NADPH-generating systems in bacteria and archaea. *Frontiers in Microbiology*, *6*(JUL), 1–27. https://doi.org/10.3389/fmicb.2015.00742

Tsuruta, H., Paddon, C. J., Eng, D., Lenihan, J. R., Horning, T., Anthony, L. C., Regentin, R., Keasling, J. D., Renninger, N. S., & Newman, J. D. (2009). High-level production of amorpha-4, 11-diene, a precursor of the antimalarial agent artemisinin, in *Escherichia coli*. *PLOS One*, *4*(2), e4489. https://doi.org/10.1371/journal.pone.0004489

University Of California, Berkeley. (2004, December 30). $43 million grant from Gates Foundation brings together unique collaboration for antimalarial drug. *ScienceDaily*. Retrieved September 3, 2020 from www.sciencedaily.com/releases/2004/12/041219151820.htm

Westfall, P. J., Pitera, D. J., Lenihan, J. R., Eng, D., Woolard, F. X., Regentin, R., Horning, T., Tsuruta, H., Melis, D. J., Owens, A., Fickes, S., Diola, D., Benjamin, K. R., Keasling, J. D., Leavell, M. D., McPhee, D. J., Renninger, N. S., Newman, J. D., & Paddon, C. J. (2012). Production of amorphadiene in yeast, and its conversion to dihydroartemisinic acid, precursor to the antimalarial agent artemisinin. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(3), E111–E118. https://doi.org/10.1073/pnas.111074010

World Malaria Report (2019). Geneva: World Health Organization; 2019. Licence: CC BY-NC-SA 3.0 IGO.

Yadav, V. G., De Mey, M., Giaw Lim, C., Kumaran Ajikumar, P., & Stephanopoulos, G. (2012). The future of metabolic engineering and synthetic biology: Towards a systematic practice. *Metabolic Engineering*, *14*(3), 233–241. https://doi.org/10.1016/j.ymben.2012.02.001

Yang, C., Gao, X., Jiang, Y., Sun, B., Gao, F., & Yang, S. (2016). Synergy between methylerythritol phosphate pathway and mevalonate pathway for isoprene production in *Escherichia coli*. *Metabolic Engineering*, *37*, 79–91. https://doi.org/10.1016/j.ymben.2016.05.003

Zhou, K., Zou, R., Stephanopoulos, G., & Too, H. P. (2012). Metabolite profiling identified methylerythritol cyclodiphosphate efflux as a limiting step in microbial isoprenoid production. *PLOS One*, *7*(11), e47513.

## SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

**How to cite this article:** Patil V, Santos CNS, Ajikumar PK, Sarria S, Takors R. Balancing glucose and oxygen uptake rates to enable high amorpha-4,11-diene production in *Escherichia coli* via the methylerythritol phosphate pathway. *Biotechnology and Bioengineering*. 2021;1–13. https://doi.org/10.1002/bit.27655