Title
Applications of Synthetic Biology for Emerging Biotechnology

Author
Belcher, Michael S

Publication Date
2023

Peer reviewed|Thesis/dissertation

Applications of Synthetic Biology for Emerging Biotechnology

By

Michael S Belcher

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Plant Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Henrick Scheller, Co-chair
Professor Jay Keasling, Co-chair
Professor Patrick Shih

Fall 2022

Abstract

Applications of Synthetic Biology for Emerging Biotechnology

by

Michael S Belcher

Doctor of Philosophy in Plant Biology

University of California, Berkeley

Professor Henrik Scheller, Co-chair

Professor Jay Keasling, Co-chair

As government and corporate policies transition to meet the UNs sustainable development goals as outlined in Agenda 2030[1], there will be a drastic need for the invention and refinement of "green technologies" applicable in a variety of economic sectors and disciplines. Biotechnology has become heavily influenced by the field of synthetic biology[2], which offers the capacity to deconstruct, shape, and rebuild natural systems from all domains of life. This is achieved via the development and integration of novel synthetic systems with natural biological processes. The potential of synthetic biology is now being realized as demonstrated by its application in numerous industries including (but not limited to): medicine[3], agriculture[4], food[5], energy[6] (both bio and petroleum based), natural product discovery and production[7], pharmaceutical drug development[8], and materials science[9].

Tools such as CRISPR-Cas9[10], along with the development of next generation DNA synthesis[11], assembly[12], and sequencing technologies[13], have unlocked the potential of synthetic biology. This has made the engineering and augmentation of most living systems possible, allowing for the development of complex and refined genetically modified organisms (GMOs) for deployment in numerous applications. The world is transitioning into the era of the Fourth Industrial Revolution[14], an era focused on the mitigation of climate change and the race to Net Zero[15]. I envision that synthetic biology will play a crucial role in this transition, and while this space is much too vast for one person to explore in totality, scientists continue to work on independent components while exploring collaboration for the synergistic application of their discoveries. Eventually, the input from varying specialties form functional "high-level" systems developed with synthetic biology.

In particular, the fields of synthetic plant biology and synthetic yeast biology have shown great promise for the development of breakthrough biotechnology platforms. The engineering of plants that serve as primary feedstocks for biofuel production (which generally has focused on maximizing the feedstock-to-fuel conversion efficiency through cell wall engineering) is now exploring the augmentation of current agricultural systems for the bioproduction of high-value biologics and therapeutics *in planta*. While still nascent in application, there is much promise for these technologies due to the scalability of *in planta* production, without the need for sterile conditions or complex manufacturing controls. The development of more efficient and versatile

methods for the engineering of synthetic genetic circuits in plants is crucial for the deployment of the highly specialized plant chassis in biotechnological settings.

Yeast, on the other hand, can serve as a counterpart to large-scale bioproduction in plants by providing a chassis for biochemical pathway discovery and complete biosynthesis of complex molecules from all domains of life. Yeast is a highly dynamic single-celled eukaryotic organism that has been highly characterized and is easily manipulated/engineered in the lab. In recent decades yeast has proven to be an effective platform for the bioproduction of various natural, specialty, and commodity chemicals whose manufacturing is not possible or cost effective with current methods. Additionally, yeast as a bioproduction platform offers the prospective of new-to-nature molecules through the coalescing of biosynthetic enzymes from disparate pathways. This is a truly exciting prospect for the future of drug discovery and development as we utilize the vast genetic diversity of the biological world for the construction of novel biosynthetic enzymes and pathways.

The following sections aim to highlight my research focused on synthetic plant and yeast biology, with a focus on the development of technologies and strategies for application in the bioenergy and bioproduction sectors. While this alone will not answer all the existential problems of a transitioning world, it represents a small piece of a very large puzzle that we as a collective are working to solve.

Applications of Synthetic Biology for Emerging Biotechnology

"Science is replacing evolution by natural selection
with evolution by intelligent design"
-*Yuval Noah Harari*

Dedication

I would like to express my deepest gratitude to my trio of mentors and advisors Dr. Jay Keasling, Dr. Henrik Scheller, and Dr. Patrick Shih that provided me with the support to scientifically explore my diverse interests. Their patience, leadership, and feedback over the last five years has been invaluable to my success and growth as a budding scientist. Even through the global challenges of 2020-2022, they were always striving to ensure I had everything required to meet the goals set forth. I am also grateful to both the UC Berkeley Graduate Division and the National Science Foundation for the financial support provided by The Berkeley Fellowship and Graduate Research Fellowship Program. The five years of funding provided by these fellowships allowed me to diversify my research and explore projects that would have not been possible otherwise. Additionally, I could not have undertaken this journey without the support of my amazing wife, who moved away from her family to join me during this pursuit of higher education. Her love and advice were crucial to getting through many of the tough times. I would also like to extend my sincere appreciation to my friends and lab mates for their assistance with planning and executing experiments. Specifically, Dr. Graham Hudson, Dr. Yuzhong Liu-Dierks, Dr. Andy Zhou, and Dr. Mitch Thompson, who were all vital to various aspects of my work, and importantly maintaining my sanity (which is still questionable). Lastly, I would like to thank all the people at The Joint Bioenergy Institute who work behind the scenes to keep the lab running for all of us, especially Mary Armitage, Sadiki Showers, and Cliff Ng for their daily contributions to my work.

Table of Contents

Table of Contents (cont.)

Figures

Acknowledgements

I would like to start by acknowledging all the brilliant scientific minds I have had the pleasure of working with over the past five years. I would not have been possible to have successes without the numerous conversations and collaborations with you all. Specifically, I would like to acknowledge the co-authors of Biofuels for a Sustainable Future published in *Cell* in 2021: Dr. Yuzhong Liu, Dr. Pablo Cruz-Morales, Dr. Amin Zargar, Dr. Bo Pang, Elias Englund, Dr. Qingyun Dan, Kevin Yin, and Dr. Jay Keasling. I would also like to acknowledge the co-authors of Design of Orthogonal Regulatory Systems for Modulating Gene Expression in Plants published in *Nature Chemical Biology* in 2020: Khanh Vuu, Dr. Andy Zhou, Nasim Mansoori, Amanda Agosto Ramos, Dr. Mitch Thompson, Dr. Dominque Loque, Dr. Henrik Scheller, and Dr. Patrick Shih. Additionally, thank you to Kasey Markel and Dr. Patrick Shih for their co-authorship of our publication Defining and Engineering Bioenergy Plant Feedstock Ideotypes as well as Jessica Mahinthakumar and Dr. Jay Keasling for their co-authorship of New Frontiers: Harnessing Pivotal Advances in Microbial Engineering for the Biosynthesis of Plant-Derived Terpenoids, both published in *Current Opinion in Biotechnology* in 2020. These published manuscripts were used when constructing my dissertation, as cited in the Table of Contents. Lastly, I would like to acknowledge all my collaborators that assisted with varying aspects of my dissertation that were unpublished at the time of writing: Jack Liu of Stanford University (Sattely lab) for his collaboration on deciphering the T5aH product spectrum, Dr. Yuzhong Liu for providing various engineered yeast lines for analysis and visualization of enzyme expression and localization dynamics, and Dr. Graham Hudson for his numerous contributions to Taxol-associated P450 localization in yeast and the biosynthesis of steroidal alkaloids in yeast.

Curriculum Vitae

Michael S. Belcher

Belcher.michael45@gmail.com

EDUCATION

University of California Berkeley — Berkeley, CA
Ph.D, *Synthetic Biology and Bioengineering* — GPA: 4.00
December 2022

Oregon State University — Corvallis, OR
Bachelor of Science, *Botany and Plant Pathology* — GPA: 3.89
Summa Cum Laude Honors — June 2017

Portland Community College — Portland, OR
Courses in: *Calculus, Biology, Chemistry,* — GPA: 3.81
Organic Chemistry, and *Physics* — 2014-2015

RESEARCH EXPERIENCE

University of California Berkeley, Plant and Microbial Biology Department — Berkeley, CA
PhD student, researcher at the Joint Bioenergy Institute (JBEI, Emeryville, CA) — August 2017-
Advisors: Dr. Jay Keasling, Dr. Henrik Scheller, and Dr. Patrick Shih — December 2022

- Development of synthetic orthogonal regulatory systems for modulating transgene expression in plants.
- Engineering cell wall composition of biofuel crops (*S. bicolor*, *P. virgatum*) to increase saccharification efficiency.
- Heterologous production and biochemical pathway characterization of novel/high-value bioproducts in plants.
- Application of tissue culture methodology for plant transformation and gene expression analysis.
- Monosaccharide and cell wall analysis of *S. bicolor*, *S. lycopersicum*, *P. virgatum*, and *N. benthamiana*.
- Design, development, and implementation of synthetic gene circuits in plants and yeast.
- *In vivo* analysis of heterologous protein dynamics with fluorescent reporter fusions and confocal microscopy.
- Development of hybrid enzymes with enhanced biochemical activity and altered subcellular localization.
- Engineering *S. cerevisiae* as a bioproduction chassis for medically relevant plant natural products.
- Characterization and biochemical evaluation of plant natural product biosynthetic pathways in *S. cerevisiae.*
- Heterologous production of saponins, taxanes, steroidal alkaloids, and cannabinoids in *S. cerevisiae.*

Oregon State University, Department of Botany & Plant Pathology — Corvallis, OR
Undergraduate Research Assistant — August 2015-August 2017
Advisor: Dr. Jeff Chang

- Investigation of a novel bacterial biocontrol of *Agrobacterium* and *Rhodococcus.*
- Characterizing the Type II Secretion System (T2SS) for roles in *Agrobacterium rhizogenes* virulence.
- Characterizing the microbiome of crown galls from various plants and geographic regions.
- Isolation, sequencing, and characterization of novel Ti plasmid types.
- Characterization of atypical plant pathogenic *Agrobacterium.*
- *Agrobacterium* genomics and diagnostics.

Portland Community College, Department of Biology — Portland, OR
Community College Undergraduate Research Initiative (CCURI) — August 2014- June 2015
Advisor: Prof. Josephine Pino

- Protoplast Isolation and Regeneration of Local Moss Species.

FELLOWSHIPS & SCHOLARSHIPS

National Science Foundation Graduate Research Fellow — Berkeley, CA
National Science Foundation (NSF GRFP recipient) — 2019-2022

Berkeley Fellow, The Berkeley Fellowship — Berkeley, CA
UC Berkeley Department of Graduate Studies — 2017-2019

Phi Kappa Phi Graduate Fellow — Corvallis, OR
Phi Kappa Phi, Oregon State University Chapter — 2017

Wayne and Joann Chambers Scholarship — Corvallis, OR
College of Agricultural Science, Oregon State University — 2016-2017

Thomas C. Moore Memorial Scholarship — Corvallis, OR
Department of Botany and Plant Pathology, Oregon State University — 2016-2017

Undergraduate Research, Innovation, Scholarship and Creativity (URISC) — Corvallis, OR
Scholarship Recipient, Division of Undergraduate Studies, Oregon State University — 2016-2017

AWARDS & HONORS

Outstanding Graduate Student Instructor- PMB department — December 2019

- Graduate student instructor for PMB 104L with Dr. Kathleen Ryan

Excellence in Teaching Award- Berkeley Graduate Division — May 2019

- Graduate student instructor for Bio 1AL

Best Poster Presentation- JBEI annual retreat — August 2019
Travel award as an invited speaker for 38th Crown Gall Conference — August 2017
Power Top-Off award: University of California Berkeley, Berkeley, CA — January 2017
Waldo-Cummings Outstanding Student Award: Oregon State University, Corvallis, OR — May 2017
Drucilla Shepard Smith Scholastic Award (4.0 Institutional GPA) — May 2016
Dean's List (>3.5 GPA): Oregon State University, Corvallis, OR — Fall 2015-Spring 2017
The Honor Society of Phi Kappa Phi: OSU Chapter, Corvallis, OR — 2016
Golden Key International Honour Society — 2016
Phi Theta Kappa Honor Society — 2015
President's List (4.0 GPA): Portland Community College, Portland, OR — Winter 2014-Spring 2015

PUBLICATIONS

Design of orthogonal regulatory systems for modulating gene expression in plants. **Belcher, Michael S**., Khanh M. Vuu, Andy Zhou, Nasim Mansoori, Amanda Agosto Ramos, Mitchell G. Thompson, Henrik V. Scheller, Dominique Loqué, and Patrick M. Shih. ***Nature Chemical Biology*** 16, no. 8 (2020): 857-865.

New frontiers: harnessing pivotal advances in microbial engineering for the biosynthesis of plant-derived terpenoids. **Belcher, Michael S.,** Jessica Mahinthakumar, and Jay D. Keasling. ***Current opinion in biotechnology*** 65 (2020): 88-93.

Biofuels for a sustainable future. Liu, Yuzhong, Pablo Cruz-Morales, Amin Zargar, **Michael S. Belcher**, Bo Pang, Elias Englund, Qingyun Dan, Kevin Yin, and Jay D. Keasling. ***Cell*** (2021).

Lignin-based resistance to Cuscuta campestris parasitism in Heinz resistant tomato cultivars. Jhu, Min-Yao, Moran Farhi, Li Wang, Richard N. Philbrook, **Michael S. Belcher**, Hokuto Nakayama, Kristina S. Zumstein et al. Recently accepted, ***Plant Physiology*** *(2022).*

Unexpected conservation and global transmission of agrobacterial virulence plasmids. Weisberg, Alexandra J., Edward W. Davis, Javier Tabima, **Michael S. Belcher**, Marilyn Miller, Chih-Horng Kuo, Joyce E. Loper, Niklaus J. Grünwald, Melodie L. Putnam, and Jeff H. Chang. ***Science*** (2020): 368, no. 6495.

An iron (II) dependent oxygenase performs the last missing step of plant lysine catabolism. Thompson, Mitchell G., Jacquelyn M. Blake-Hedges, Jose Henrique Pereira, John A. Hangasky, **Michael S. Belcher**, William M. Moore, Jesus F. Barajas et al. ***Nature communications*** 11, no. 1 (2020): 1-10.

Defining and engineering bioenergy plant feedstock ideotypes. Markel, Kasey**, Michael S. Belcher**, and Patrick M. Shih. ***Current opinion in biotechnology*** 62 (2020): 196-201.

Cell Wall Compositions of Sorghum bicolor Leaves and Roots Remain Constant Under Drought Conditions. Scavuzzo-Duggan, Tess, Nelle Varoquaux, Mary Madera, John Vogel, Jeffery Dahlberg, Robert Hutmacher, **Michael S. Belcher**, et al. ***Frontiers in plant science*** 12 (2021).

Evolutionary transitions between beneficial and phytopathogenic Rhodococcus challenge disease management. Elizabeth A. Savory, Skylar L. Fuller, Alexandra J. Weisberg, William J. Thomas, Michael I. Gordon, Danielle M. Stevens, Allison L. Creason, **Michael S. Belcher**, et al. ***Elife*** 6 (2017): e30925.

Phylogenomic analyses of the evolution of Rathayibacter toxicus and its ecological interactions with bacteriophage CS14Φ. Edward W. Davis, Javier, Tabima, Lucas Dantas Lopesa, Tal Pupkoa, **Michael S. Belcher**, Aaron J. Sechlere, Brenda K. Schroeder, Timothy D. Murray, Douglas G. Luster, William L. Schneider, Elizabeth E. Rogers, Fernando Andreote, Niklaus Grunwald, Melodie L. Putnam, and Jeff H. Chang. ***Mbio*** 9.4 (2018): e01280-18.

VOLUNTEER & WORK EXPERIENCE

iCLEM Summer Research Program
Summer 2019, 2021

- Instructor and activity designer for immersive 10-week summer research program at JBEI for local highschoolers.

Bay Area Scientists in Schools (BASIS) program
August 2017 - 2020

- Teach local elementary school kids about plant adaptations with fun and interactive games.

Communication, Literacy & Education for Agricultural Research (CLEAR)
August 2017 – 2020

- Public education on innovative agriculturally based science through open, non-confrontational communication.

STEM Camp 2016 and 2017 June 2016, June 2017

- Assist highschoolers in laboratory experiments and computational analyses to learn about biotechnology.

Math, Chemistry, and Biology Tutoring 2015-2021
Server and Bartender at Big River Restaurant, Corvallis, OR July 2015- March 2016
Server and Bartender at Oswego Grill, Lake Oswego, OR August 2012- June 2015
Server and Bartender at P.F. Chang's China Bistro, Portland, OR May 2010- August 2012
Server at Olive Garden Italian Restaurant, Escondido, CA April 2005- January 2010

Introduction

A brief analysis of the past, present, and future of biofuels

Rapid increases of energy consumption and human dependency on fossil fuels have led to the accumulation of greenhouse gases and consequently, climate change. As such, major efforts have been taken to develop, test, and adopt clean renewable fuel alternatives. Production of bioethanol and biodiesel from crops is well developed, while other feedstock resources and processes have also shown high potential to provide efficient and cost-effective alternatives, such as landfill and plastic waste conversion, algal photosynthesis, as well as electrochemical carbon fixation. In addition, the downstream microbial fermentation can be further engineered to not only increase the product yield but also expand the chemical space of biofuels through the rational design and fine-tuning of biosynthetic pathways toward the realization of "designer fuels" and diverse future applications.

There is a clear need to transition energy dependence from fossil fuels to renewable energy sources to address the unprecedented pace of climate change due to the accumulation of greenhouse gases (GHGs) in the atmosphere. Overwhelming evidence has shown that human activity is the major driver of climate change and that its consequences are impacting food production, migration patterns, economic, and political stability on a global scale. In the US alone, 6.677 gigatons of GHG were emitted in 2018 with the largest fractions being attributed to transportation (28%), electricity generation (27%), industry (22%), commercial and residential applications (12%), and agriculture (10%)[16]. As all these activities are largely dependent on fossil fuels, technological advances and diversification of alternative energy sources hold promise to significantly reduce carbon emissions and alleviate climate change.

Predating the use of petroleum itself, biofuels such as vegetable oils, animal fats, and ethanol were used for heat and illumination **(Figure 1)**. This is exemplified by the first mass-produced car, the Ford model T, which ran on corn-derived ethanol. As automobile production became increasingly industrialized in the early 20th century, it became evident that the ethanol production scheme could no longer meet the ever-growing fuel demand for internal combustion engines. Today, with environmental policies pushing for a reduction of GHG emission, aided by recent advances in crop engineering and fermentation processes, bioethanol and biodiesel production have once again become viable and sustainable surrogates for petroleum-based fuels. Bioethanol is derived from corn and sugar cane in the United States and Brazil, respectively, which together account for 84% of the total global production. In the United States, bioethanol production has reached a volume of 15.7 billion gallons in 2019[17], thus meeting the mandatory 10% supplementation requirement for gasoline[18]. In Europe, the lack of cultivable land and the ban on genetically modified crops has largely limited bioethanol production. As such, 75% of the biofuel market in the European Union is composed of biodiesels derived from rapeseed, palm oil, soybean, and used cooking oil. As of 2015, biofuels have reduced carbon emissions by 589.3 million tons **(Figure 1)** and will continue to play an important role in renewable energies[19].

Recent advances in battery technology have substantially increased the power density of electrical energy storage devices, thus accelerating the development of electric vehicles. However, to date, electricity in the US is still predominantly derived from fossil fuels such as gas and coal (38% and 23%, respectively)[20]. Furthermore, limitations in the driving range, high capital cost, the lack of infrastructure, and power-to-weight ratios preclude the implementation of

electric long-haul vehicles and aviation. To reach a carbon-neutral to -negative transportation scheme, a more diversified approach therefore requires the use of both electric vehicles and biofuels alike. Specifically, electric vehicles hold promise in short-range and light-weight configurations, whereas the use of biofuels offers significant advantages for conventional long-distance ground transportation and aircraft.

To mitigate GHG emissions while meeting the global fuel demand, biofuel technology advancements need to focus on (1) optimization of current biofuel-production technology for higher productivity and efficiency of lignocellulosic biomass conversion, (2) diversification of feedstocks to ensure the viability of biofuel production within existing ecological and economic constraints (e.g., carbon fixation through photosynthetic and electrochemical means as well as conversion of biowaste into value-added products), and (3) expansion of the chemical space toward designer molecules that improve fuel economy and performance while reducing carbon emissions. Major efforts need to be devoted not only to overcome technological barriers but also to integrate social, economic, and environmental factors to provide long-term, cost-effective, and reliable production systems for the biofuel industry.

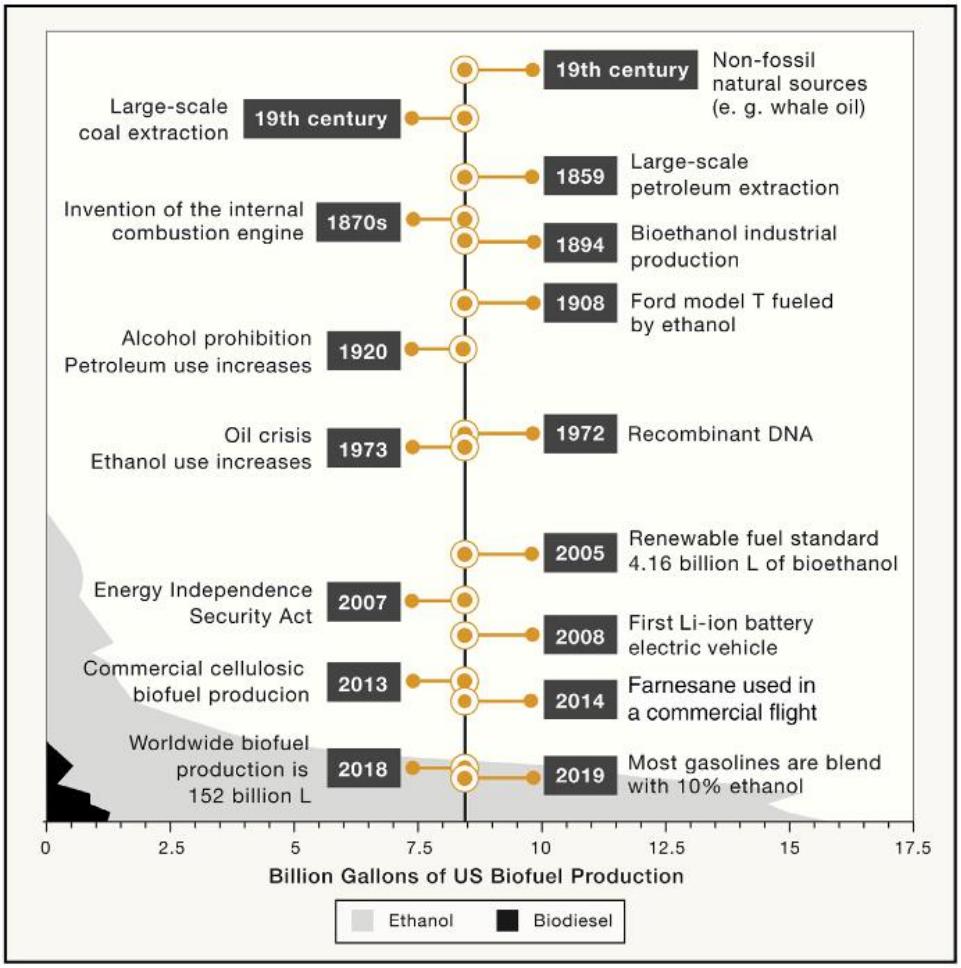


Figure 1. A timeline of biofuel production. Non-fossil fuels have been used long before the beginning of the oil era in the late 19th century. Scientific advancements, legislation, and the pressing need to lower CO 2 emissions has led to an increase in the production of biofuels, reaching 15.8 billion gallons and 1.7 billion gallons per year of ethanol and biodiesel, respectively.

Advancing lignocellulosic feedstocks for biofuel production

Biofuel technology has evolved through several generations of significant advancements. The predominant problem with first-generation biofuels is that they are derived from food crops (e.g., corn and sugar cane), which require fertilization, water, and soil, and thus directly compete with food production. Tight regulations on the use of pesticides and genetically modified crops further limit their utilization in sustainable transportation. In order to mitigate these short-comings, second-generation biofuels are derived from non-edible lignocellulose remnants of plants, which consist of up to 70% polymerized sugars and constitute the most abundant form of biomass on Earth[21]. These biofuels are attractive because their net carbon footprint (emitted carbon – consumed carbon) can be neutral or even negative, and their generation from agricultural and forest residues or white wood chips provides economic advantages compared to crops[22,23]. However, using lignocellulose for biofuel production requires energetically and financially costly extraction of fermentable sugars such as thermal, chemical, and/or biochemical pre-treatment. As a result, despite the fact that the Energy Independence and Security Act (EISA) of 2007 set an annual blending target of 16 billion gallons of cellulosic biofuels by 2022 for the US[24], by 2017 production had amounted to less than 2% of this benchmark[17]. Significant technological progress has since been made in the production of lignocellulosic biofuel toward a clean and economically viable process, including advances in energy crop engineering strategies, efficient degradation of lignocellulose, and simultaneous manufacturing of higher-value products.

The climate benefits of large-scale lignocellulosic biofuel production were initially questioned due to its potential competition with land use for reforestation; it is believed that energy crop cultivation may result in less carbon capture efficiency than reforestation, leading to a carbon debt that must be compensated by the carbon negativity of the resulting biofuels. However, recent analysis of switchgrass production on transitioning crop/pasture land showed that in fact, its GHG mitigation potential is comparable with reforestation of this land and has several times more mitigation potential than grassland restoration (Field et al., 2020)[22]. Additionally, the ability of energy crops such as sorghum to grow on marginal lands provides an avenue for biofuel production that minimizes the competition for necessary farmable land to support the growing population[25,26].

In order to maximize the land use for lignocellulosic biofuel production, crops have been engineered to be more productive in accumulating biomass by increasing their photosynthetic capacity and carbon fixation efficiency **(Figure 2)**. Biological processes like non-photochemical quenching (NPQ) and photorespiration are bioconversion of photon energy into fixed carbon. This is because the NPQ process dissipates excess photon energy as heat (unproductive), and the transition from an NPQ state to a carbon fixation state (productive) is generally slow, leading to mass energy loss in field conditions. It has been shown that overexpressing the genes responsible for NPQ relaxation in the model crop *Nicotiana tabacum* can accelerate the switching process, resulting in ~15% increases in plant height, leaf area, and total biomass accumulation[27]. Additionally, plants have evolved to maximize light capture with much of the energy wasted. While counterintuitive, diminishing a plant's light harvesting capacity in dense field conditions has a drastic and beneficial effect on biomass accumulation. In fact, truncation of light-harvesting complex antenna components decreased the capacity for light capture in engineered lines resulting in a 20% increase in total biomass accumulation under these conditions[28]. Photorespiration is another process that limits productivity due to Ribulose-1,5-Bisphosphate

Carboxylase/Oxygenase's (RuBisCO's) capacity to react with molecular oxygen in place of CO2 thus leading to a net loss of carbon sequestration efficiency. Therefore, introducing an engineered photorespiratory bypass pathway into biofuel crops has the potential to increase the energy conversion efficiency of the plant[29,30]. Eventually, many of these traits can be stacked into individual biofuel crops to increase plant productivity and maximize biomass production as technologies for advanced plant engineering and gene regulation are developed[31].

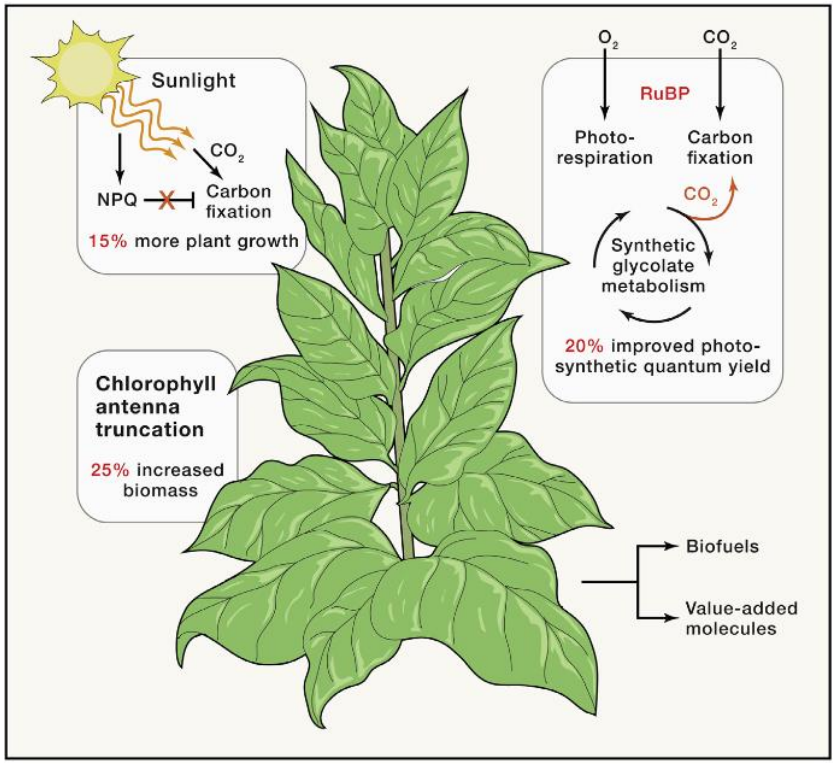


Figure 2. Engineering strategies in planta to improve CO 2 fixation. Acceleration of NPQ relaxation, truncation of the chlorophyll antenna, and introduction of a photo respiratory bypass are promising strategies to increase CO 2 fixation and biomass yield in plants.

Metabolic pathways to synthesize products with higher values can also be incorporated in the meantime to increase the economic feasibility of lignocellulosic biofuels as fossil fuels are currently produced at a far lower price than biofuels, and simultaneous manufacturing of higher-value products with biofuels would increase their economic viability. ''Molecular farming,'' which couples agriculture with the production of high-value small molecules and proteins including therapeutics and antibodies, is a promising strategy for increasing the value of lignocellulosic biomass (Buyel, 2019; Yang et al., 2020)[32,33]. Accumulating these molecules after integrating their biosynthetic pathways in a biofuel crop background and introducing efficient extraction schemes into the processing pipeline can drastically decrease production costs, thus increasing profits. This has additional implications for medicine, as cancer biologics and viral antibodies can be produced in planta at high levels in the field without the need for sterile manufacturing systems (Capell et al., 2020;Dent and Matoba, 2020; Donini and Marusic, 2019; Mortimer, 2019)[34,35,36,37]. This is one of the most promising strategies to achieve economic viability for biofuel production, although transgene biocontainment strategies will need to be implemented to prevent unwanted transgene flow from engineered crops (Clark and Maselko,

2020)[38]. Because of the mismatch in the volumes of fuel needed versus the volume of each individual therapeutic needed, it will be necessary to have a large number of crops, each producing the same biofuel precursor and different high-value products, which will be agronomically challenging.

Biofuel production from lignocellulosic biomass relies on the microbial bioconversion of cell-wall sugars and components into fuels and products (Baral et al., 2019; Perez-Pimienta et al., 2019)[39,40]. A major hurdle to efficient bioconversion is the recalcitrance of the feedstock material and the inhibitory effect that lignin has on this process (Dos Santos et al., 2019)[41]. Cell-wall engineering has shown promise for decreasing overall recalcitrance by increasing the ratio of C6/C5 sugars, reducing lignin content, and reducing the acetylation of cell-wall polymers that limit the conversion efficiency of the feedstock material (Aznar et al., 2018; Eudes et al., 2015; Yan et al., 2018)[42,43,44]. While lignin is a major contributor of feedstock recalcitrance, it is also a promising substrate for specialized microbes that convert these aromatic polymers into usable products (Fang et al., 2020; Incha et al., 2019)[45,46]. The introduction of specialized microbial hosts into various processing systems has the potential to optimize the conversion of all lignocellulosic feedstock components into products with economic value, limiting the waste streams for biofuel production and increasing the viability for their use on a global scale.

The synergistic application of these various strategies has the potential to make lignocellulosic biofuels economically viable while shifting the current paradigm of what an effective biofuel/bioproduct production system achieves. Through a multidisciplinary approach across all sectors, we have the potential to revolutionize the manufacturing of biofuels/bioproducts from lignocellulosic biomass ushering in a new era of green technologies.

Diversifying photosynthetic feedstocks for biofuel production

While the first and second generations of biofuels use light and CO2 to produce biomass in crops that is later fed to microbes, third-generation or algal biofuels combine energy capture and fuel production within a single cell of photosynthetic cyanobacteria and algae **(Figure 3)**. Having the entire fuel-production process take place in one organism makes the process more direct and efficient with no energy invested in non-fermentable parts such as plant stems, roots, and leaves. The solar energy conversion in cyanobacteria and algae is higher than that in plants, reaching an efficiency of 3% in microalgae compared to less than 1% in most crops (Wijffels and Barbosa, 2010)[47]. Furthermore, many species can grow in wastewater or marine environments with simple nutritional requirements and therefore do not compete for land use with agriculture. It is estimated that microalgae can produce oil at a yield of 100,000 L/hectare/year, while palm and sunflower oil can only reach 1,000–6,000 L/hectare/year. Algal fermentation could also lead to 9,000 L/hectare/year of bioethanol production, compared to 600 L/hectare/year derived from corn (Alalwan et al., 2019)[48].

Despite these favorable comparisons, attempts at large-scale cultivations have struggled with high production costs. Unlike agriculture, which has been optimized over millennia by humans, the technology for mass scale cultivation of photosynthetic microorganisms is still in its early developmental stage. The cultivation can be done in either an open system like a raceway pond, or in a closed system such as a photobioreactor. Open ponds have lower operating costs, but their use is limited due to the risk of contamination, as well as strict regulation against genetically modified organisms in an uncontained system (Abdullah et al., 2019)[49]. Closed systems, on the

other hand, can have more tightly controlled cultivation conditions and have a low risk of contamination but the operating costs are high: $2,743 per ton of biomass versus $1,227 per ton of biomass in open ponds (algae farm cost model) (Zhu et al., 2018)[50]. These costs can be lowered if the following limitations are addressed: (1) light dissipation in the NPQ process, (2) narrow usable light spectrum, and (3) poor carbon fixation efficiency of microbes.

Light dissipation in NPQ in algae is akin to that of energy crops, which can also be addressed by a similar engineering strategy: truncation of the chlorophyll A in green algae (Melis et al., 1999)[51]and modification of the light-harvesting complex antenna size of the cyanobacterial photosystems have been shown to improve the solar-to-product conversion efficiency by up to 3-fold (Melis, 2009)[52]. The second issue is cyanobacteria and microalgae only capture the visible light range (400–700 nm), which makes ∼50% of incident solar energy inaccessible (Blankenship et al., 2011)[53]. However, terrestrial cyanobacteria, when grown in far-red light environments, express a novel chlorophyll f (Chen et al., 2010)[54]. Heterologous expression of this pigment in *Synechococcus sp.* PCC7002 successfully extended light absorption up to 750 nm, thereby broadening the usable solar spectrum for fuel production (Tros et al., 2020)[55].

The issue with poor carbon fixation is more complex in nature since RuBisCO, the main CO2 fixing enzyme, suffers from poor catalytic activity and its promiscuity toward O2. In fact, attempts to improve the activity of RuBisCO have had limited success due to a trade-off between CO2 affinity and carboxylation rate (Savir et al., 2010)[56]. As an alternative, efforts to engineer the Calvin cycle for increased regeneration of RuBisCO's substrate, ribulose-1,5-biphosphate (RuBP), resulted in a 69% increase in ethanol yield (Liang et al., 2018)[57]. In addition, CO2 fixation can be improved by engineering a carbon sink that pulls fixed carbon away from the Calvin cycle. By introducing the 2,3-butanediol pathway into *Synechococcus elongatus* PCC 7942, as well as enzymes generating pyruvate from 3-phosphoglycerate, the total carbon yield was increased by 1.8-fold (Oliver and Atsumi, 2015)[58]. Unnatural carbon fixation pathways have also been introduced to circumvent the poor properties of RuBisCO and its complicated regulation. Certain archaea use a reductive acetyl-CoA pathway to fix carbon with formate dehydrogenase and CO dehydrogenase/acetyl-CoA synthase. This pathway is ATP efficient, requiring 2 ATPs to synthesize acetyl-CoA, compared to 7 in the Calvin cycle (Claassens et al., 2016)[59]. However, due to the pathway normally operating under anaerobic conditions, it might have limited transferability to organisms with oxygenic photosynthesis. Nevertheless, an in vitro study has demonstrated that enoyl-CoA carboxylases/reductases with better properties than RuBisCO, along with 16 other enzymes from all three domains of life can be incorporated in a circular pathway where cofactors can be regenerated (Schwander et al., 2016)[60].

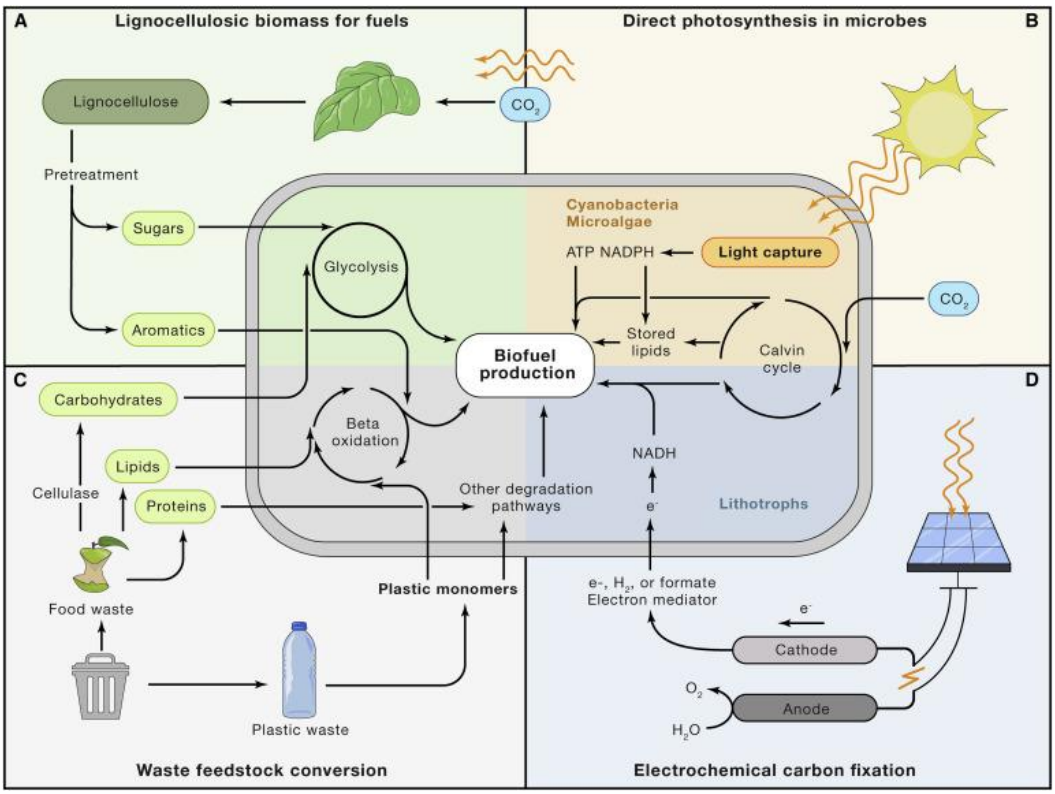


Figure 3. Current diversity of feedstock conversion pipelines for biofuel production. (A and C) Upon pretreatment, sugars and aromatic molecules can be extracted from lignocellulosic biomass; together with carbohydrates, lipids, proteins, and plastic monomers from waste, these molecules can be used as carbon sources for the microbial production through glycolysis and beta oxidation pathways to make biofuels and biogas. (B) Instead of feeding biomass to microbes, photosynthesis and direct synthesis of biofuels can be achieved in a single cell in cyanobacteria and microalgae. (D) Lithotrophs can be coupled to the cathode of an electrochemical cell, with delivery of electrons from the electrode driving CO2 reduction and carbon fixation.

One key parameter that dictates the viability of algal biofuels is the productivity of the host strain. Certain microalgae can store up to 80% of their dry weight as lipids, making them an attractive target for biodiesel production, while other strains accumulate carbohydrates that can be fermented to make ethanol (Alalwan et al., 2019)[61]. Further engineering strategies have demonstrated the feasibility to upregulate lipid production (e.g., triacylglycerol, De Bhowmick et al., 2018)[62] and convert residual proteins and carbohydrates toward further fermentation (Rashid et al., 2013)[63]. The lipid length and level of saturation can also be altered through metabolic engineering, which can yield biodiesel for direct use in vehicles without extensive engine modification (Kings et al., 2017)[64]. To produce a specific biofuel directly by photosynthetic microorganisms, metabolic engineering is often required, and cyanobacteria are typically used due to their genetic tractability. A range of different fuel molecules have been produced, including alcohols, free fatty acids, molecular hydrogen, and alkanes (Knoot et al., 2018)[65]. Standout examples include 5.5 g/L (212 mg/L/day) of ethanol and 4.8 g/L (302 mg/L/day) of butanol, both in the cyanobacterial model strain *Synechocystis sp.* PCC6803 (Gao et al., 2012;

Liu et al., 2019)[66,67]. In these cases, the most efficient strategy to reach high titers was to use the strongest promoters available to drive overexpression of pathway enzymes. However, a better understanding of the native metabolism and the development of more sophisticated genetic tools will likely lead to more efficient approaches. Indeed, transcriptomics studies have revealed the importance of small RNAs for cellular control (Kopf et al., 2014)[68], and genome scale models have allowed coupling production with growth (Shabestary and Hudson, 2016)[69]. Still, the field has a long road ahead before cyanobacteria can reach the productivity of model heterotrophic hosts.

Defining and engineering bioenergy plant feedstock ideotypes

Ideotypes are theoretical archetypes of crops which serve as a practical framework for plant breeders to critically evaluate what traits they should be targeting for specific applications. With advances in plant biotechnology and a growing urgency to adopt more sustainable practices across our economy, new uses for crops as bioenergy feedstocks may pivot our definition of an ideal crop that is engineered for biomass and bioenergy production, in contrast to food production. Although there is a plethora of specific applications to which plant engineering efforts can contribute, here we highlight recent advances in two broad areas of research: increasing available plant biomass and engineering production of higher value co-products.

Before our ability to genetically engineer plants, plant breeders were constrained to breeding and selecting from the morphological, physiological, and metabolic repertoire already preexisting in plant genomes. Initially, such efforts were focused on breeding out deleterious traits or on a narrow aim such as yield. Fifty years ago, the concept of an ideotype was proposed as an alternative regime. The ideotype is an idealized form of a particular crop, which could then be a target to breed towards, rather than merely breeding away from deleterious traits[70]. This shift in mentality provided a much-needed framework to help set goals and target traits for plant breeding efforts. A useful ideotype must be 'theoretically capable of greater production than the genotype it is to replace and of such design as to offer reasonable prospect that it can be bred from the material available[70]. The discovery and development of plant genetic engineering technologies such as *Agrobacterium*-mediated and biolistic transformation expanded the scope of possible ideotypes, as plant engineering efforts can now draw on a much larger effective pool of genetic material, expanding from interfertile germplasm to all sequenced and characterized genes from across the tree of life. Feedstock crops are harvested primarily for biomass, which is then used as a substrate for downstream processes (e.g., bioconversion, fermentation, combustion, etc.). Thus, it becomes useful to frame plant carbon partitioning in terms of biomass composition, and what production or deposition of small molecules or polymers would be present in feedstock ideotypes. Using new synthetic biology tools to redesign carbon flow in plants, one may alter and optimize the composition of biomass and bioproducts in a way that cannot be achieved through conventional breeding methods, ultimately improving the scalability and feasibility of renewable feedstock crops. The ideotype for each crop may vary depending on its economics, growing region, and intended application. We propose some aspects of possible ideotypes for several biomass crops. Here, we focus on carbon allocation as a metabolic/ physiological trait that may be modified to increase the utility and value of feedstock crops. Specifically, we focus on two aspects: 1) traits that may alter overall plant biomass and the usability of this biomass and 2) traits that may enhance the value of feedstock crops with the

production of higher value co-products, paying special attention to advances within the last two years.

Engineered traits to enhance plant biomass

The plant cell wall is a complex network of polymers and is one of the most effective carbon sequestering systems on the planet, with annual production of land plants estimated at 150–170 billion metric tons per year[71]. Cell walls represent a massive and largely untapped supply of six-carbon (C6) sugars in the form of cellulose (b-1,4-linked glucose). However, cell walls are naturally recalcitrant to degradation and fermentation, limiting their use as chemical feedstocks rather than bulk materials[72]. Lignin is a main inhibitor of saccharification (sugar release) in woody crops and hemicellulose limits saccharification yields in monocot biomass crops[73]. Many engineering efforts have focused on decreasing lignin and improving fermentation characteristics. We are only beginning to explore ways to modify the composition and deposition of plant cell wall components to improve their ability to serve as biomass feedstocks. One strategy for reducing lignin accumulation uses 3-dehydroshikimate dehydratase (QsuB) from *Corynebacterium glutamicum*, which converts a lignin precursor into protocatechuate. Transgenic expression of *QsuB* in Arabidopsis thaliana plastids reduced lignin accumulation and improved saccharification yield by 25-100% depending on treatment method[74]. Moreover, the six-carbon/five-carbon (C6/C5) sugar ratio of the biomass also affects saccharification yields, with higher ratios performing better. The most highly accumulated C5 sugar is xylose, but xylan (xylose polymer) synthesis mutants show dwarfism due to xylem vessel collapse. This phenotype has been rescued by returning xylan synthesis specifically to vessel tissue, leading to a 42% increase in saccharification yield compared to wild type[75]. Acetylated cell wall components are converted during fermentation to acetic acid, which inhibits fermentation. RNA-interference (RNAi) has been used to decrease expression of genes responsible for acetylation, nearly tripling saccharification yields[76]. Gene stacking has been used to generate engineered lines that contain multiple aforementioned traits[77]. This demonstrates how modern bioengineering strategies can be used in tandem to modify the cell wall composition, a step towards engineering the optimum bioenergy crop ideotype. While ideotype specifics will vary by crop and intended application, in general an idealized biomass cell wall will have a high C6/C5 sugar ratio, low lignin concentration, and provide a favorable substrate for fermentation.

Beyond modifying the molecular composition of the cell wall, others have also focused on engineering upstream metabolic processes to increase rates of photosynthesis, carbon fixation, and biomass production. Plants often absorb more photons than they can use for photosynthesis, leading to non-photochemical quenching (NPQ) that dissipates excess energy as heat but does not contribute to biomass. Mutation of light harvesting complex components results resulted in a 25% biomass increase in *Nicotiana tabacum* under field conditions[78]. It is also possible to modulate the NPQ process to shift more quickly from a heat-producing to a photosynthetic state, restoring energy capture via production of NADPH and ATP. Engineered N. tabacum overexpressing the genes coordinating NPQ relaxation showed increases of ~15% in plant height, leaf area, and total biomass accumulation in field conditions[79]. These are promising results, as most plants use similar mechanisms making this technology applicable to bioenergy crops dependent on the maximum accumulation of lignocellulosic biomass. Another key process that limits the theoretical maximum for biomass accumulation is photorespiration. The primary cost of photorespiration stems from the process plants use to ‘recycle’ the unintended product

formed via the oxygenase activity of RuBiSCO, leading to loss of both carbon and nitrogen. An alternative photorespiratory bypass based on the 3-hydroxypropionate bicycle was successfully engineered into cyanobacteria by expressing six heterologous genes from *Chloroflexus aurantiacus.* This bypass not only limits losses from photorespiration, it also fixes additional carbon and can supplement the Calvin-Benson cycle[80]. Other photorespiratory bypasses have been demonstrated to work in planta yielding more than a 25% increase in biomass in field trials[81]. Thus, the ability to modify both the rate of carbon fixation and the fate of carbon deposition in the form of various cell wall polymers have been shown to be complementary processes for increasing the accessible feedstock sugars from future feedstock plant crops.

Engineered co-products to increase feedstock value

Lignocellulosic bioproduction offers a much larger potential supply of biomass than food-based fuels such as corn-ethanol, and reduces the conflict between food and fuels, materials, and other products which may be produced from biomass crops. Future biomass crop ideotypes should therefore be designed to ensure the use of lignocellulosic material is cost effective. Lignocellulosic biofuels have been slow to achieve commercial viability, in part due to low fuel prices and the chemical recalcitrance of lignocellulosic matter[82,83]. A promising strategy to make lignocellulosic biofuels economically competitive is the co-production of higher value products directly in feedstock crops, which can be separated from the bulk carbon fuel source during processing[84]. This can be achieved in two ways: either feedstocks for lignocellulosic biofuels can be modified so as to produce a higher value side product, or lignocellulosic biofuel can be produced from side products of other agricultural processes. The former is amenable to feedstock bioengineering efforts to optimize for biofuel purposes and will be discussed here. The ideotype of co-product crops will depend on the specific crop, but one important component is that the co-product sells for more than the cost of extraction. Co-product value (in terms of $/kg) and market size tend are often inversely correlated, as shown in **Figure 4**.The base use of most biomass crops is production of ethanol, but plants have been engineered to produce co-products such as higher value fuels, commodity chemicals, and high value small molecules. Higher value fuel products include lipids for biodiesel and jet fuel. Biodiesel-grade lipids have recently been produced in engineered sorghum that accumulates 8% dry weight oil in leaves in the form of lipid droplets[85]. These droplets can be extracted using simple, cheap techniques during the standard processing pipeline for lignocellulosic biofuels, minimizing additional purification costs. Jet fuel is also a high-volume product with an annual market size of 290 billion liters in 2015, with prices usually ranging around $1 per liter. There is no practical alternative available for liquid aviation fuels[86], which account for a small but rapidly growing fraction of total anthropogenic greenhouse gas emissions- currently 2.3% and growing at approximately 6% per year[87]. Jet fuels have been produced from the oilseed crop camelina, and efforts are underway to increase jet fuel yield[88]. Another promising high-volume side product is 1,5-pentanediol, a commodity chemical used in polyester and polyurethane production. The present market value is around $6000/ton, with a market size of 18 million USD[84]. Using plants as a production chassis for high value low volume products has received substantial attention in recent years, with several analyses suggesting plants may allow for cheaper production of edible vaccines, bulk enzymes, and monoclonal antibodies than alternative systems[89]. These high value products split

into two major classes: high value small molecules and proteins. The anti-malarial drug Artemisinin is a high value small molecule naturally produced in nonmodel plants.

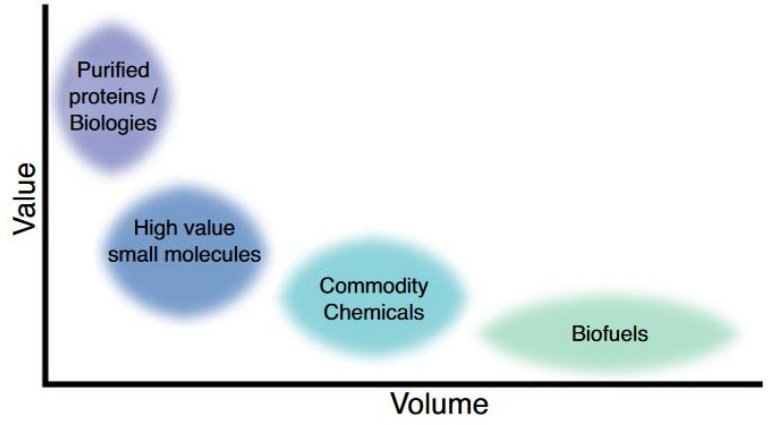


Figure 4. Tradeoff between value and volume for engineered co-products. Inverse correlation between product volume and value per unit, adapted from[98].

Fast-growing tobacco plants have recently been engineered to produce the drug, potentially expanding production capacity[90]. Plants have been the primary production platform for high value small molecules for millennia - medicines, spices, and drugs have mostly been sourced from plant hosts. In modern times, the large-scale production of cannabis and opium poppies attest to the scalability and cost-efficacy of in planta small molecule production. Despite advances in microbial engineering and synthetic chemistry, plants remain the production platforms of choice for these high value small molecules, demonstrating the low cost and high scale that can be achieved with plant systems. With an appropriate ideotype compromising between fermentable sugars and high-value co-products, engineered lignocellulosic biomass crops may be economically viable.

Rather than a single ideotype for all biomass crops, different crops may be more amenable hosts to particular applications. Complex metabolic pathways to produce high value small molecules have been successfully implemented in model plant species, and some biomass crops seem particularly amenable to metabolic engineering for high value small molecule production as well as overall modifications to the total carbon allocation **(Figure 5)**. Ultimately, engineered feedstock crops that produce co-products may help offset costs associated with a future plant-based bioeconomy that will have to compete with petrochemicals.

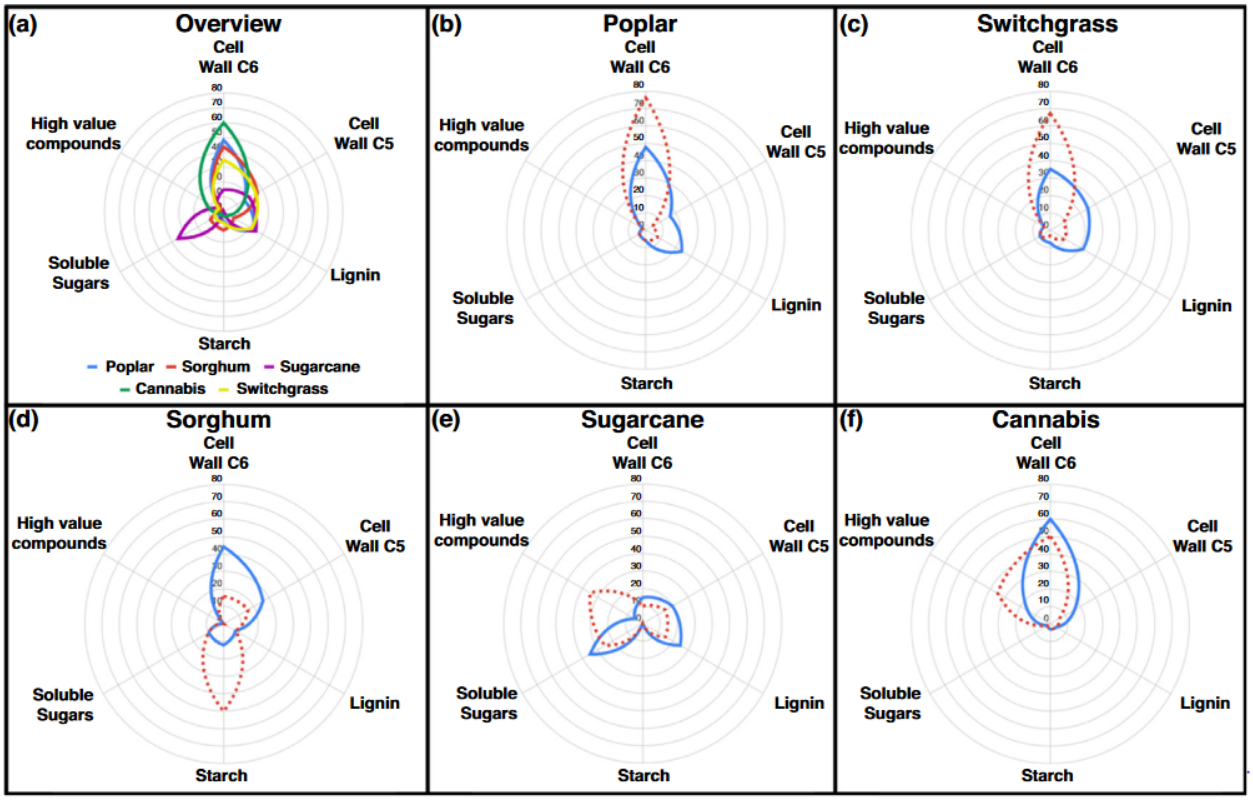


Figure 5. Carbon allocation schemes for various biofuel crops of interest. Carbon allocation in five bioenergy and bioproduct crops. Spider-plots demonstrating carbon allocation between six major categories: cell wall C6, cell wall C5, lignin, starch, soluble sugars, and high value small molecules. Panel A shows best estimates of current carbon allocation between these six categories. For this panel, blue indicates poplar, red indicates sorghum, purple indicates sugarcane, green indicates cannabis, and orange indicates switchgrass. Panels B–F show current allocation (solid blue line) and a hypothetical ideotype (dashed red line) for the crops: poplar, switchgrass, sorghum, sugarcane, and cannabis, respectively. Concentric rings are percent dry weight, the outermost circle for all charts is 80% to allow for comparison between panels. Data are averages from literature sources and should be taken as approximates. See **Table 1** for data sources.

	Poplar	Switchgrass	Sorghum	Sugarcane	Cannabis
Cell wall C6	45% [30]	37% [31]	36% [32]	43% [33]	48% [34]
Cell wall C5	16% [30]	25% [31]	23% [32]	19% [33]	13% [34]
Lignin	26% [30]	16% [31]	16% [32]	26% [33]	3% [35]
Starch	2% [36]	5% [37]	12% [38]	<2% [39]	3% [40]
Soluble sugars	3% [36]	10% [41]	9% [42]	35% [43]	2% [44]
Small molecules	N/A	N/A	N/A	N/A	3–30% [45]

Table 1. Percentage of carbon allocation in potential feedstock crops. All data are presented as percentage dry weight and rounded to the nearest percent. When data are tissue-specific, averages or data from the most abundant tissue were used. All data are from untreated, wild-type, or control plants. (Adjusted citations for Table 1 are 99-114 for source data.)

Conclusion

Dedicated crops have been used in first-generation food-to-ethanol production for over 100 years [91], and in the United States annual production has increased 10-fold since 1990[22]. Ethanol accounts for over 90% of all biofuel produced in the United States, nearly all of which is derived from dedicated fields of corn, consuming 38% of corn production[92]. The production of biofuel products from food crops causes competition between food and fuel, raising the price of staple foodstuffs[93,94]. Lignocellulosic 'second generation' biofuels substantially reduce this problem by either growing on marginal land where food crops are not viable, or by production from agricultural residues rather than diverting a food crop into the biofuel pathway.

Biofuels have sometimes been presented as an environmentally friendly and low-carbon alternative to fossil fuels, but current implementations have failed to deliver substantial GHG emission reductions. Biofuels grown from established agricultural fields generally achieve GHG emission reductions of 20–80% compared to fossil fuels[95]. However, land use change associated with the conversion of natural land to biofuel production leads to a 'carbon debt' that takes decades to centuries to pay back, negating any GHG savings[96]. Furthermore, conversion of natural land to biofuel production is a major driver of rainforest loss[97]. Growing biofuel feedstock on marginal lands and producing biofuel as one of multiple products are the two main strategies to reduce this trade-off. Here we consider re-designing biofuel feedstock crops to reduce cell wall recalcitrance, increase biomass per acre, and generate additional products not only add value but also improve resource use efficiency.

Modern biotechnology has expanded the possibilities of crop ideotypes by allowing for plant phenotypes not attainable through classical breeding. Petrochemical fuels have been instrumental for global industrialization, and their use remains indispensable at present. However, climate considerations as well as the practical limitations inherent in using a finite resource call for the development of alternative sources of liquid fuel and materials. Plant biomass is the most viable means of production sufficiently scalable to take the place of petrochemicals in the economy of the future, and ideotype breeding serves as a useful paradigm for the design and improvement of biomass feedstock crops.

Applications of synthetic plant biology for bioengineering

No matter the goal of a plant engineering project, gene expression and the regulation of its expression is crucial for synthetic biology applications. At its core, the regulation of when, where, and how strong a gene is expressed underpins the phenomenon of multicellular life. Intricate networks of transcriptional elements (transcription factors, promoters, RNAs, enhancers, and regulatory proteins to name a few) harness complex arrangements of "logic gates" to conduct the development of a multi-system organism from a single embryonic cell. The development of synthetic regulatory systems that provide control of when a gene is expressed (specific to an external condition or developmental stage), where it is expressed (which tissue or cell type), and the level of expression (regulate accumulation of mRNA and protein products) without disrupting endogenous processes would allow for the engineering of complex, multilayered, synthetic gene circuits in future endeavors. In the work presented here, we sought to develop novel tools for the modulation of transgene expression in plants by developing synthetic orthogonal regulatory elements. These elements allow for the refinement of the engineering scheme by providing the means to tune the desired synthetic genetic circuit with a higher level of precision. We also developed a series of synthetic genetic circuits building upon our preliminary work with synthetic transcriptional regulators for deployment in sorghum.

Design of orthogonal regulatory systems for modulating gene expression in plants

Agricultural biotechnology strategies often require the precise regulation of multiple genes to effectively modify complex plant traits. However, most efforts are hindered by a lack of characterized tools that allow for reliable and targeted expression of transgenes. We have successfully engineered a library of synthetic transcriptional regulators that modulate expression strength *in planta*. By leveraging orthogonal regulatory systems from *Saccharomyces spp.*, we have developed a strategy for the design of synthetic activators, synthetic repressors, and synthetic promoters and have validated their utility in *Nicotiana benthamiana* and *Arabidopsis thaliana.* This characterization of contributing genetic elements that dictate gene expression represents a foundation for the rational design of refined synthetic regulators. Our findings demonstrate that these tools provide variation in transcriptional output while enabling the concerted expression of multiple genes in a tissue-specific and environmentally responsive manner, providing a basis for generating complex genetic circuits that process endogenous and environmental stimuli.

Plants offer a unique platform to address many imminent challenges that face society, as future engineering efforts hold promise in promoting sustainable agriculture, renewable energy, and green technologies[115]. However, the tools to effectively modify and engineer plants are still in their infancy. One major hurdle has been the development of genetic parts that enable precise control of transgene expression in plants. Many genetic and metabolic engineering strategies require robust and accurate control of multiple genes to optimize synthetic pathways, regulate flux, and introduce new traits[116,117]. The ability to modulate gene expression provides a direct approach to address these tasks. However, the majority of plant engineering efforts are limited to a small number of characterized constitutive promoters, which may result in unintended pleiotropic effects or toxicity issues and are limited in their range of expression strength.

This work was inspired by previous studies that have successfully engineered orthogonal gene expression tools employing a reductionist and modular approach to parts design[118-123]. Our unique strategy builds upon this approach by blending elements from both yeast and plants and could theoretically be applied to any TF type. Additionally, many of these previous systems were constructed using repetitive *cis*-element DNA sequences in the promoter design, while an important aspect of our approach is maximizing sequence diversity while tuning promoter strength. Specifically, we sought to build an expansive and diverse library of synthetic transcriptional regulators for plant engineering with components from various yeast transcription factor (TF) systems in conjunction with plant-specific regulatory DNA sequences. With this approach we developed a method that can be expanded beyond the well-characterized Gal4-based synthetic systems that have been used in the past[124-126]. In general, parts design was approached in three ways: altering TF DNA-binding dynamics at the regulatory promoter by introducing DNA *cis*-element variation with randomly concatenated *cis*-elements, modulating RNA polymerase II recruitment by testing various plant minimal promoters, and directly modifying TFs from disparate families through truncation and fusion of activation or repression domains to generate new synthetic TFs, or *trans*-elements **(Figure 6).**

A major finding of our work is that plant minimal promoter sequences can be leveraged for the design of chimeric promoters. By utilizing *cis*-elements from yeast in combination with a plant minimal promoter, we can generate functional synthetic promoters that interact with both the orthologous yeast TF as well as the endogenous transcriptional machinery of the plant. The minimal promoter, also known as a core promoter, lies directly upstream of the transcription start site and is where the transcription pre-initiation complex including RNA-polymerase II binds[127,128]. Thus, the minimal promoter region facilitates the assembly and stabilization of the pre-initiation complex leading to variable levels of basal and activated transcription determined by this interaction[129,130]. The binding of additional TFs, specifically orthogonal *trans*-elements in this work, at the regulatory promoter upstream of the minimal promoter can stimulate or repress this basal level of transcription providing multiple layers of regulation[131]. Using this strategy, we characterized a diverse set of synthetic promoters and synthetic *trans*-elements composed of discrete genetic parts from various TF families. Promoters were constructed with a collection of TF-binding *cis*-elements appended upstream of a plant minimal promoter. All minimal promoter sequences also contain the native 5' UTR downstream of the transcription start site. We tuned promoter strength by modifying basal transcription through minimal promoter variation and modulating the binding dynamics of the *trans*-element with concatenated *cis*-element (CCE) variation. The initial library of synthetic promoters was developed for the yeast TF, Gal4, by appending random combinations of five well-characterized Gal4 upstream activation sequences (i.e., *cis*-elements), to varied plant minimal promoters. Through our approach, we explored how these CCEs and minimal promoters contribute to the overall strength of a synthetic promoter. We then expanded upon this design strategy to TFs from disparate protein families and assembled a full suite of *trans*-elements and synthetic promoters for each. By expanding our library to contain both *cis*- and *trans*-element variation, we generated and tested more than 500 unique promoter/TF pairs resulting in a wide range of transcriptional output potential. Our library of TF-binding *cis*-elements, plant minimal promoters, and TF fusion proteins demonstrates a novel method for the design, construction, and characterization of new tools for the controlled modulation of gene expression for various plant synthetic biology applications.

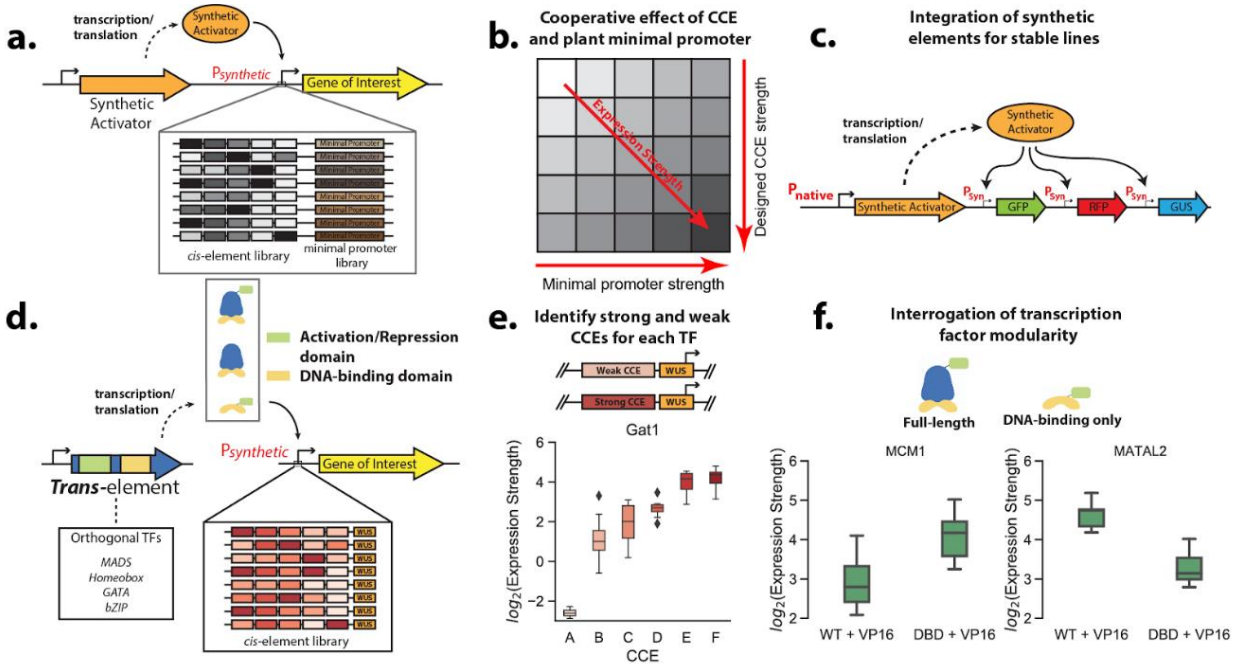


Figure 6. Visual schematic of design principles and experimental workflow.
a, A randomized library of CCEs and minimal promoters was used to generate a suite of synthetic promoters for the Gal4 expression system to identify contributions of each element to the strength of a synthetic promoter. **b**, A combinatorial library of five select CCEs and minimal promoters was generated for Gal4. We found that CCEs and minimal promoters have an additive contribution to gene expression. **c**, A single endogenous promoter, denoted as pNative , was used to drive the expression of an orthogonal synthetic activator in a tissue specific or environmentally responsive manner. This synthetic activator then induced expression of three reporter genes to verify reporter presence corresponds with the tissue or condition tested**. d**, The effect of trans-element and CCE diversity on gene expression was examined using numerous TF families from S. cerevisiae. Increasing the diversity of TF families tested allows for an investigation into how our modifications behave in varying systems and expands our parts library drastically. **e**, Testing a single trans-activator with the corresponding suite of synthetic promoters verifies that the CCE has a direct effect on gene expression strength for the other TF families. This variation provides the capacity to control the relative expression strengths of multiple genes when coupled with a single trans-element and increases the sequence diversity of our parts. The center line of the box and whisker plot denotes the median value, the box represents the interquartile range while whiskers represent the 5th and 95th percentile with values outside considered outliers ($n \geq 8$ biological replicates). **f**, The modularity of our trans-element design was investigated by fusing activation domains to the minimal DNA binding scaffold of each TF. We found that the fusion of an activation domain to the truncated TF, denoted as DBD, had varying effects on transcriptional output dependent on the family used for the basis of the design. The center line of the box and whisker plot denotes the median value, the box represents the interquartile range while whiskers represent the 5th and 95th percentile with values outside considered outliers ($n \geq 8$ biological replicates).

Sequence randomization modulates expression strength

As a proof of concept for our promoter design strategy, we utilized the well-characterized TF Gal4 fused to the VP16 activation domain as the activating *trans*-element[132]. The heterologous nature of yeast TFs like Gal4 provides an opportunity to leverage a purely orthogonal system in plants, decoupling the transcriptional regulation of transgenes from those endogenous to the genome. To test the abundance of parts generated in this study, we used a high throughput transient expression assay in *Nicotiana benthamiana*. This system allows for the rapid screening of parts using a combinatorial approach, with each element cloned into individual binary vectors.

A diversity of known Gal4 *cis*-elements were manually curated from native promoters in the yeast Gal regulon, each with a distinct nucleotide sequence and assumed to exhibit a diversity of dissociation constants with Gal4[133]. We hypothesized that these deviations would lead to variation in the transcriptional output of each synthetic promoter, depending on the combination and position of *cis*-elements used in the CCE design. Minimal promoters from a collection of plant promoter sequences were appended to the CCEs generated with these Gal4 *cis*-elements to produce complete synthetic promoters **(Figure 7a)**. This strategy also presents an opportunity to investigate the effect the minimal promoter has on gene expression outside the context of its endogenous sequence[134,135]. Additionally, introducing nucleotide diversity to our promoter constructs may limit the potential for transcriptional silencing often observed when identical sequences are used multiple times in gene stacking efforts[136-139]. We screened these randomized combinations to measure the strength of the promoter in the presence of our synthetic *trans*-element and the basal expression of our promoters in the absence of the TF. A measurement of Green Fluorescent Protein (GFP) fluorescence was used as the proxy for transcriptional output, while the constitutive expression of a Red Fluorescent Protein (RFP) was used as a normalization metric, with the ratio of GFP over RFP providing normalized values for the output of each construct. As expected, we observed a distribution of expression strengths while avoiding the usage of identical sequences, demonstrating a strategy for tuning gene expression **(Figure 7c)**.

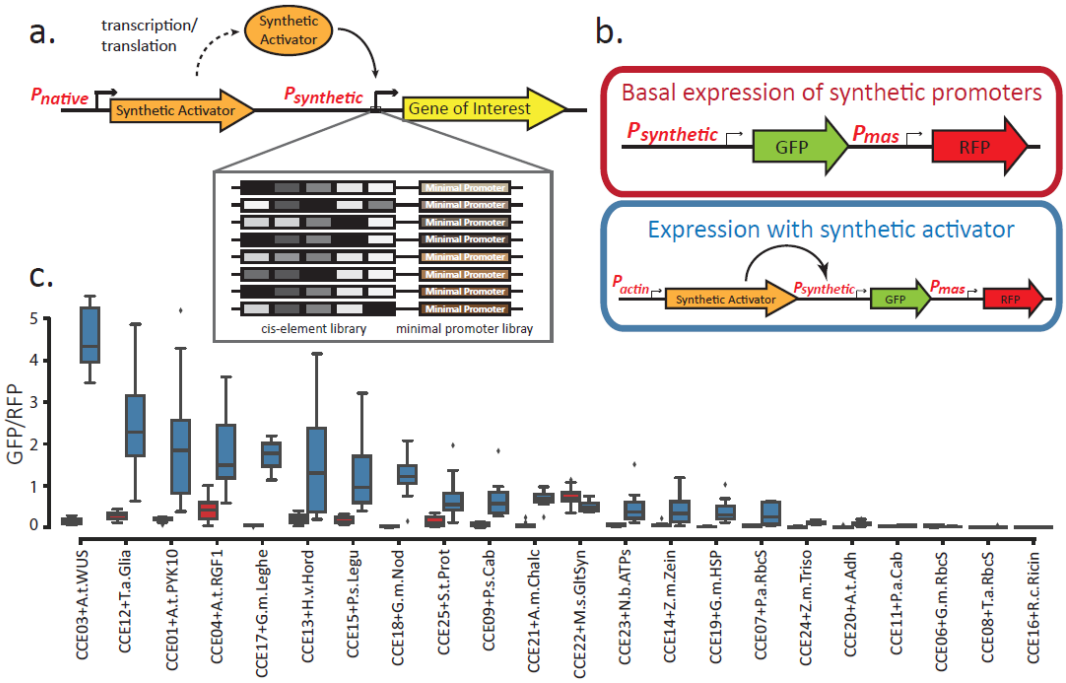


Figure 7: Design and characterization of a library of synthetic promoters.
a, Brute force strategy to design and generate a library of promoters with varying expression strengths. The synthetic activator was synthesized by fusing a Gal4 DNA-binding domain to a nuclear localization sequence and VP16 activator domain. A library of *cis*-elements that bind Gal4 and vary in sequence were gathered from endogenous yeast promoters that fall within the Gal regulon. Various plant minimal promoters were also gathered. Five random *cis*-elements were concatenated to generate CCEs and subsequently added upstream to a plant minimal promoter to generate synthetic promoters. **b,** Synthetic promoter strength was characterized by their capacity to drive GFP expression. The constitutive *MAS* promoter from *Agrobacterium* was used to drive RFP, downstream and in the same orientation, to normalize between samples. The synthetic activator was driven by the constitutive *ACT2* promoter from *A. thaliana*, enabling the expression of GFP. A control construct was generated, lacking the synthetic activator, to measure the basal expression of the synthetic promoters. **c,** A range of expression strengths can be observed with the designed synthetic promoters. Constructs including the synthetic activator enable GFP expression (blue), while controls lacking synthetic activators provide basal expression levels of synthetic promoters (red). Constructs were transiently expressed in *N. benthamiana* leaves, and GFP fluorescence was normalized to constitutive expression of RFP and reported in arbitrary units. The center line of the box and whisker plot denotes the median value, the box represents the interquartile range while whiskers represent the 5th and 95th percentile with values outside considered outliers ($n \geq 8$ biological replicates)

Promoter parts independently augment expression strength

Our library of randomly designed synthetic promoters provides the opportunity to initiate rudimentary investigations into the genetic components of plant promoter architecture. The comparatively simple and better-characterized promoter structure of prokaryotes has enabled the design of various synthetic promoter systems[140,141]. However, even in the most well studied eukaryotic systems (*e.g.*, yeast), we are only beginning to elucidate the genetic components that

dictate gene expression[142]. To test the contributions of various CCEs and minimal promoters, we constructed a combinatorial library of five CCE sequences and five minimal promoters based on performance in the initial screen. The expression strength of all twenty-five promoters was measured, and the expression output revealed each element has an independent effect on promoter performance **(Figure 8)**. That is, specific CCEs were identified that drive higher gene expression by altering the binding of the *trans*-element, and similarly, minimal promoters were identified that increase expression strength in both the basal and induced states. The empirical characterization of this library demonstrates the ability to engineer synthetic promoters with tunable expression determined by pre-characterized parts.

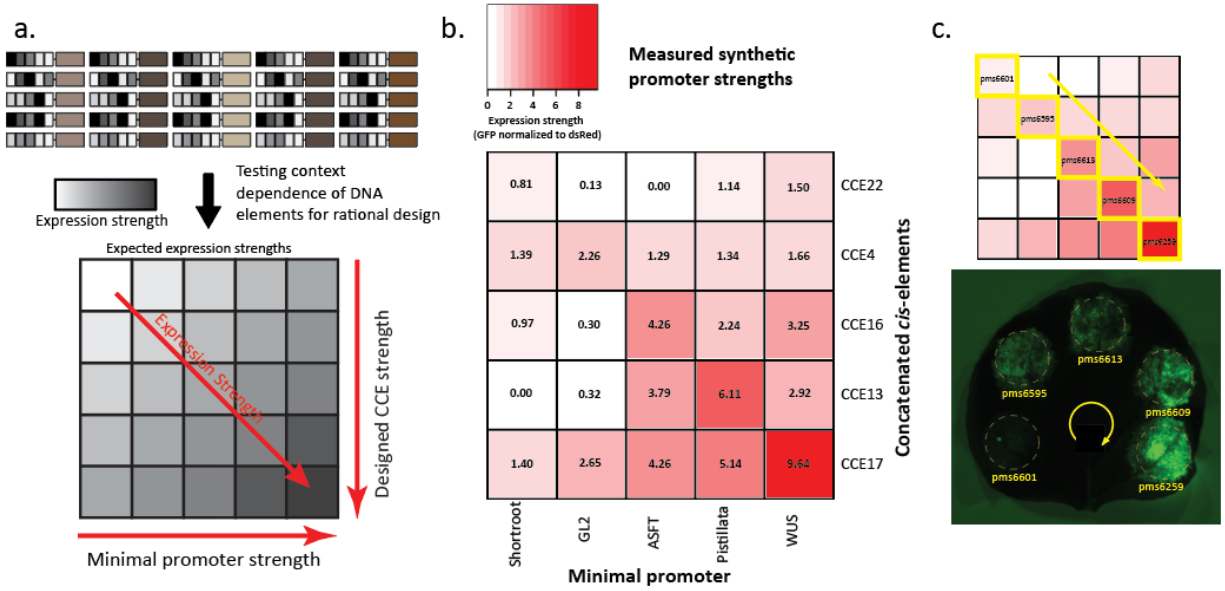


Figure 8. Teasing apart the contribution of DNA elements reveals generalizable trends in promoter expression strength. a, Combinatorial design of CCE and minimal promoter to test the independent contributions to expression strength in a predictable manner. Promoter strengths were expected to increase with the addition of a strong CCE and minimal promoter, whereas weak expression would be expected when using a weak CCE and minimal promoter. **b,** Characterization of all combinations of synthetic promoter generated from five CCEs and five minimal promoters spanning variable expression strengths. Parts were selected based on data from Figure 1, with the goal of selecting elements with varying strength. Promoter output is measured as GFP fluorescence normalized to constitutively expressed RFP, in the same construct design described in Figure 1. The general trend was strong output being correlated with the usage of DNA elements that promote expression; however, some noise is still observed, which can be due to the context dependence of promoter elements within the genome. All values represent the average of $n \geq 8$ biological replicates. **c,** Qualitative demonstration of five synthetic promoters. This displays the trend of increasing promoter strength with the usage of DNA elements that promote higher expression. Five constructs (labeled and outlined) with varying expression strengths were infiltrated into *N. benthamiana* and imaged.

Synthetic parts enable state-specific gene expression

In many cases, it is not necessary to constitutively express all transgenes, and thus state-specific regulation of gene expression provides an additional dimension of control over a genetic circuit. For example, the constitutive overexpression of a given protein may act as a sink on cellular resources resulting in overall detrimental effects[143]. Similarly, various agricultural traits (*e.g.*, disease resistance) often result in fitness costs, and thus targeted expression of these genes may curtail unintended consequences[144,145]. To further evaluate the efficacy of our parts *in planta*, we generated stable lines in *A. thaliana.* To demonstrate that these synthetic elements can coordinate the expression of multiple genes in a tissue-specific manner, we expressed the synthetic activator under an endosperm-specific promoter, *At2S3*, stacked with three reporters (GFP, RFP, and beta-glucuronidase (GUS)) driven by unique synthetic promoters **(Figure 9a)**. As the reporters can only be turned on by the synthetic activator, we observed the expected expression of all three exclusively in seed tissue **(Figure 9b)**. Additionally, examination of the vegetative tissue arising from the seeds with active reporters showed no reporter activity, with expression cycling from seed to seedling and from flowering plant to seed. Our findings demonstrate the use of synthetic promoters for the tissue-specific regulation of multiple genes simultaneously.

Coordinating the expression of multiple genes in an inducible or environmentally responsive manner may provide a solution to many challenges in plant engineering. To highlight how our synthetic promoters may address some of these issues, we designed a similar circuit with the synthetic activator driven by a phosphate responsive promoter, *AtPht1.1*, which is induced under low external phosphate concentrations. Stable *Arabidopsis* transformants display the expression of all reporter genes in response to phosphate deprivation in the medium, and as expected, reporters were not observed when phosphate was supplemented **(Figure 9c)**. These results display how our synthetic system may enable control over stacked genes in applications that require engineering integrated with environmental cues (*i.e.*, abiotic and biotic stresses). Importantly, these data also demonstrate our system can be transferred between different species *(i.e., N. benthamiana* and *A. thaliana*) while retaining its functional properties. Additionally, no phenotypic or growth disturbances were observed demonstrating the orthogonality of the parts. This strategy can be further expanded in future studies by linking this cue to a series of transcriptional events by expressing a second (third, fourth, fifth, etc.) *trans*-element under a synthetic promoter to generate multiplexed transcriptional cascades originating from a single endogenous signal.

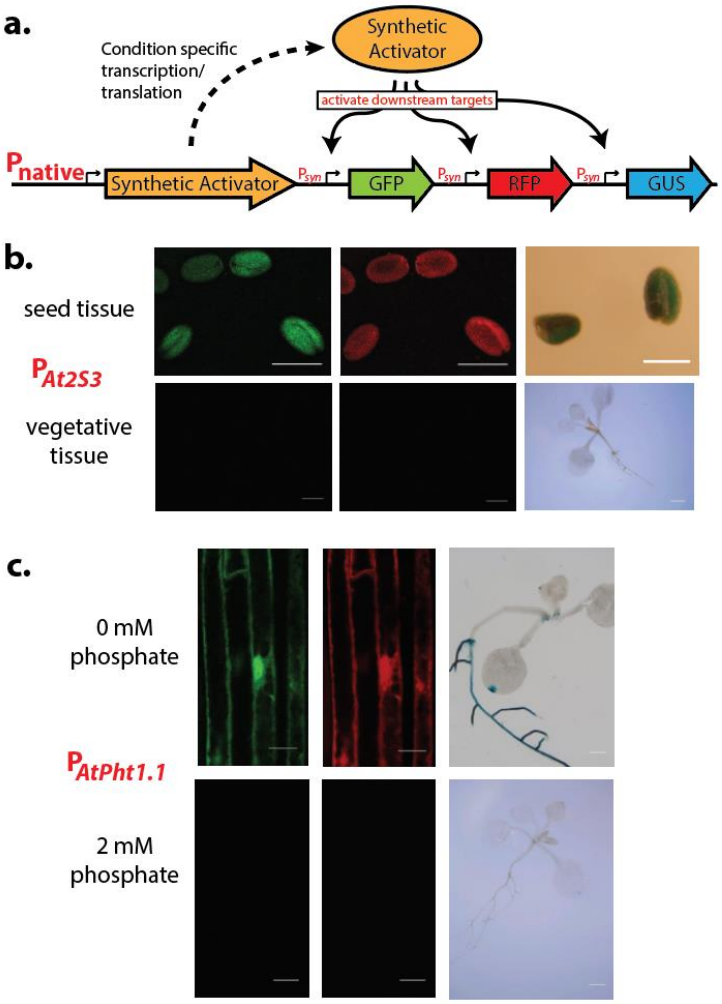


Figure 9. Utilizing synthetic promoters for the coordinated expression of multiple stacked transgenes. a, Schematic of how a synthetic activator can be utilized to drive the concerted expression of multiple downstream genes of interest. The cartoon depicts the basic design of constructs used to demonstrate the expression of multiple transgenes (GFP, RFP, and GUS) can be controlled by regulation of the synthetic activator. CCE03+A.t.WUS was used to drive GFP expression, CCE01+A.t.PYK10 was used to drive RFP, and CCE02+A.t.PRK was used to drive GUS. **b,** Spatial regulation of multiple reporter genes using endosperm-specific expression of the synthetic activator driven by the *At2S3* promoter. Seeds from transgenic plants show expression of all three reporter genes, whereas vegetative tissue taken from roots or whole seedlings showed no indication of reporter gene expression. Scale bar for seed images measures 0.5mm, scale bar for vegetative tissue is 1.0cm **c,** Regulation of multiple reporter genes under the control of the synthetic activator which responds to environmental stimuli. The *AtPht1.1* promoter is driving the synthetic activator, enabling the inducible expression of all three downstream reporter transgenes in response to phosphate deprivation. Scale bar for root cell images is 40µm, scale bar for seedling with 0mM phosphate is 0.5cm, scale bar for seedling with 2mM phosphate is 1.0cm.

Exploring trans-element modularity with disparate TFs

Although Gal4-based synthetic TFs have previously been developed for plant systems, there are clear applications that would benefit from the expansion and development of other synthetic TFs. After validating our promoter design strategy, we investigated how conducive other TFs from different protein families would behave in our system, specifically MADS (MCM1), Homeobox (Mata1 and Matα2), GATA (Gat1), and bZIP (Yap1) type TFs. We tested how these TFs behave when truncated to the predicted DNA-binding domain, and their potential to be functionally reconstructed through the fusion of transcriptional regulatory domains **(Figure 10a)**. Thus, in an effort to expand our parts library and explore the versatility of our approach, we designed and characterized additional sets of synthetic TFs along with corresponding promoter libraries. *Cis*-

elements for Yap1 and Gat1 were designed using experimentally determined sequence motifs based on position-specific affinity matrices with randomly chosen bases for ambiguous nucleotides in each motif[146]. *Cis*-elements for MCM1, Mata1, and Matα2 were generated with sequences from previously characterized yeast promoters and were designed to bind all three TF types[147,148]**.** We subsequently generated CCEs by randomly concatenating five corresponding *cis*-elements. Complete synthetic promoter libraries for each set were then assembled through addition of the *WUS* minimal promoter, chosen for its high output in the initial library.

Our orthogonal TF design strategy introduces protein modifications to alter expression strength while expanding the parts library. We tested two known transactivation domains (TADs), C1 from *Zea mays* and VP16 from herpes simplex virus type 1, to examine their effect when appended to the tested TFs[149]. Our *trans*-element library was generated by fusing these TADs to either the full-length TF or the truncated TF consisting of solely the predicted DNA-binding domain . Additionally, these predicted DNA-binding domains were tested without a TAD fusion. We selected two predicted domains for MCM1, Gat1, and Yap1 (denoted as either D1 or D2), and a single domain for Mata1 and Matα2 (denoted as D). The prospect of designing a minimal and modular DNA-binding protein on which various activation or repression domains can be appended could drastically expand the space for synthetic tool development. All modified *trans*-elements also include an SV40 nuclear localization sequence to ensure proper import into the nucleus[150]. Using this strategy, we designed and synthesized various activating *trans*-elements for each TF and compared their expression output across a subset of their corresponding promoter library.

We designed our screen to evaluate the efficacy of transcriptional activation using truncated TFs as a minimal DNA-binding scaffold. Comparing truncated TFs to their full-length versions revealed that minimal DNA-binding scaffolds could be directly fused to TADs to enhance gene expression, and in most cases outperform the full-length TF **(Figure 10)**. The addition of a TAD to the truncated Gat1 TF chassis resulted in increased expression over the full-length TF with or without an activator, showing behavior congruent to our expectations **(Figure 10b)**. An interesting result was observed when analyzing how the DNA-binding domains of MCM1 and Yap1 behave without the addition of a TAD. Often times these *trans*-elements consisting of solely the DNA-binding domain resulted in substantial increases to expression strength. This would imply that the predicted DNA-binding domain of these constructs has inherent activation properties that may be interrupted when a TAD is fused. While the molecular basis of these observations has not been elucidated in this study, this phenomenon reveals the potential of limiting the genomic footprint of synthetic elements by utilizing minimal units with desirable transcriptional regulation properties in synthetic circuit designs. For both homeobox TFs tested (Mata1 and Matα2), the *trans*-elements generated by the addition of the C1 or VP16 activation domain to the full-length TF generally resulted in higher expression levels than those built on the minimal DNA binding scaffold. This may be due to the removal of the flexible C-terminal tail that provides stability to the TF/DNA/protein transcriptional complex[151]. While this was the observed trend, there were exceptions as demonstrated in **Figure 10e**.

Overall, the modifications we made to the native TF increased expression output, except in limited cases, as shown for Matα2 with promoter pMAlpha_9 where the full-length TF yielded higher expression than the altered versions **(Figure 10f).** Importantly, we also demonstrate how both the promoter and the *trans*-element used to drive its expression alter the overall behavior.

This highlights the importance of a thorough and systematic investigation into how *cis* and *trans* modifications modulate overall behavior. While not all parts function as expected, this work highlights a strategy for improving the development of other synthetic transcriptional systems from defined minimal biological parts and provides a characterized library for immediate use.

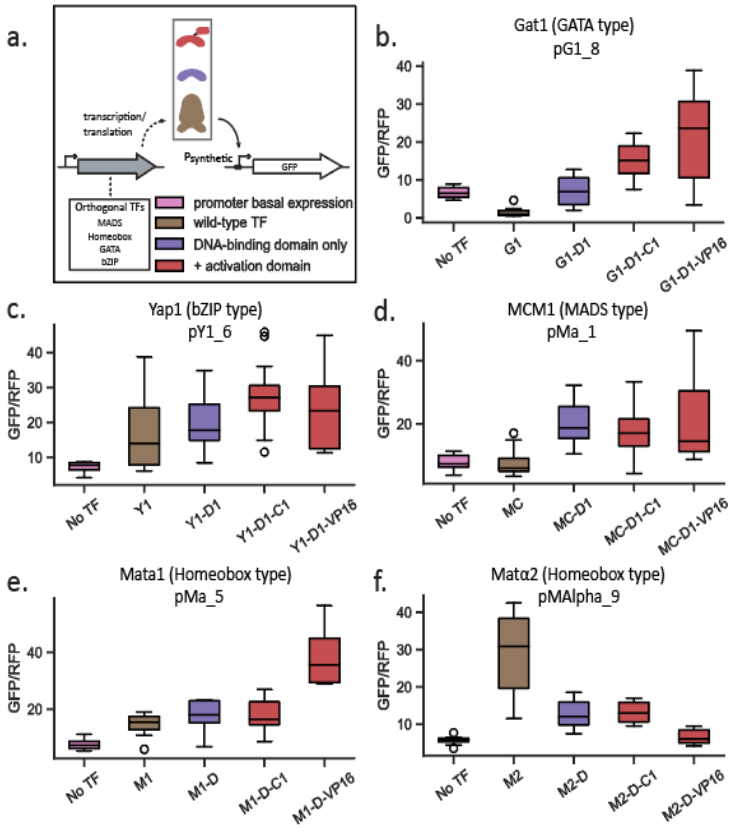


Figure 10. Fusions of TADs to both full-length and minimal DNA-binding domains provides a level of modularity to *trans*-element design. a, Graphical schematic demonstrating the *trans*-element design strategy implemented in this study. Specifically shown here is the comparison of the wild-type TF to the truncated DNA-binding domain with or without the C1 or VP16 TAD. **b-f,** Demonstrates the effect that modifications to the TF used to drive expression have on promoter output. These data display varying trends observed for each family, specific to a single promoter from the corresponding library, and confirms the potential for designing modular *trans*-elements. The center line of the box and whisker plot denotes the median value, the box represents the interquartile range while whiskers represent the 5th and 95th percentile with values outside considered outliers ($n \geq 8$ biological replicates).

Expanding TF modularity with repressor domain fusions

An additional level of control can be designed into our system with the introduction of repressive regulators, permitting logic principles in genetic circuits. To explore this potential, we designed repressive *trans*-elements that bind our synthetic promoters to repress transcription. As a proof of concept, we fused the SRDX repression domain to the Gal4 DNA-binding domain[152]. When synthetic repressors were co-infiltrated into tobacco leaves with synthetic promoters driving GFP, GFP fluorescence decreased, indicative of a repression in gene expression **(Figure 11a, 11b)**. We then generated additional repressive *trans*-elements by fusing the SRDX domain (-SRDX for C-terminal fusions and -SRDXn for N-terminal fusions) to the other orthogonal TFs we tested **(Figure 11c)**. Addition of the SRDX domain to other TF types engendered repressive *trans*-elements capable of limiting the basal expression of their corresponding promoters. Though, repression functionality was often dependent on the location of the terminal fusion, these trends varied from family to family. Plant engineers can utilize these synthetic promoters in

tandem with repressor logic to achieve tissue specific gene repression, by driving the repressing *trans*-element with a tissue-specific promoter or build complex gene circuits that can be both activated and repressed. Surprisingly, the addition of the SRDX domain to some of the *trans*-elements resulted in no change in expression, or even the increase in expression in some case. It is important to note, modifications made to a TF from one family may be consistent in behavior with the modifications made to another, as context dependence plays a role in determining functional protein fusions. Often times the addition of a given regulatory domain alters behavior inverse to the expected, highlighting the need to empirically test various modification schemes when designing parts for optimal behavior *in planta*. Nonetheless, our findings summarize a set of *trans*-elements that can be used as effective repressors for future plant engineering efforts.

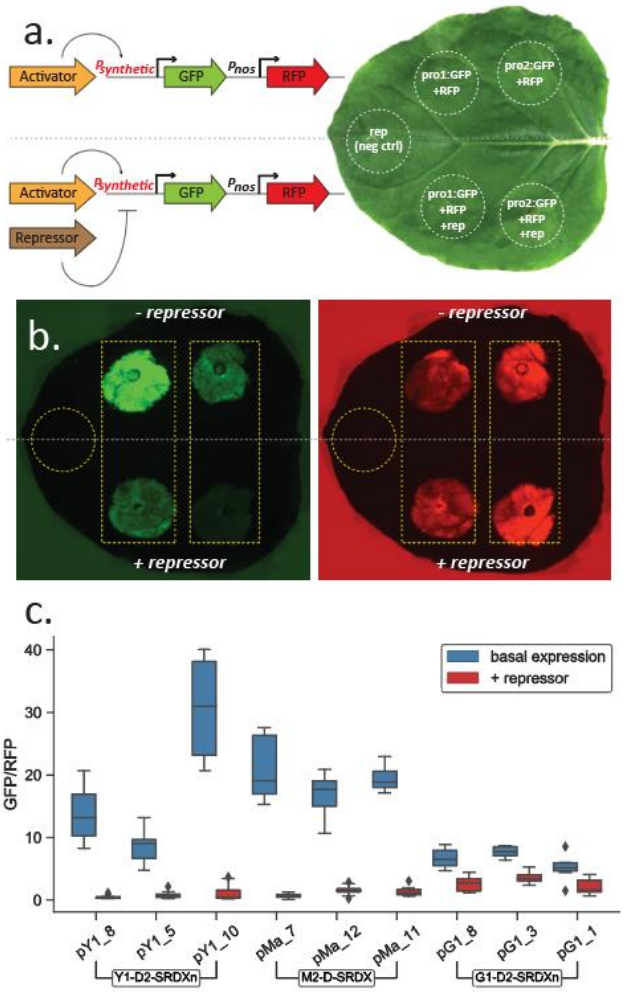


Figure 11. Utilizing synthetic repressors enables synthetic promoter compatibility with repressor logic. a, Schematic of constructs used to demonstrate how the synthetic repressor inhibits the synthetic promoter to repress transgene expression. Samples above the grey dashed line are only expressing the synthetic activator, whereas samples below the dashed line have co-expression of the synthetic repressor. Two different synthetic promoters (one high and one medium expression strength) were used to demonstrate the effects of the synthetic repressor. RFP expression is constitutively driven by the *NOS* promoter to enable normalization. Infiltration of a construct only expressing the synthetic repressor was used as a negative control. **b,** GFP expression regulated by two different synthetic promoters. Spots infiltrated without the repressor are above the dashed line, while spots below the dashed line have been co-infiltrated with the synthetic repressor. Samples within vertical dashed rectangles correspond to each other with and without the synthetic repressor. Constitutive expression of RFP allows for an internal control and normalization of GFP expression, as it is generally expressed at the same level in all samples. **c,** Quantification of repression observed with the introduction of the synthetic repressors from Yap1, Matα2, and Gat1. Functional Yap1 and Gat1 repressors were constructed with an N-terminal fusion of SRDX while the functional Matα2 required a C-terminal SRDX fusion . The center line of the box and whisker plot denotes the median value, the box represents the interquartile range while whiskers represent the 5th and 95th percentile with values outside considered outliers (n ≥ 8 biological replicates).

Hybrid promoters designed to bind multiple TF types

To expand the dynamic potential of our system, we developed synthetic promoters with *cis*-elements designed to bind multiple TF types; Mata1, Matα2, and MCM1. This was inspired by the native yeast mating system that determines haploid cell compatibility, and regulates the switch from haploid-specific gene expression to diploid-specific gene expression after mating[147,148,153-155]. These promoters are regulated by MCM1, a-specific TFs (Mata1), and α-specific TFs (Matα2) at the cis-regulatory region, with output determined by the combination bound at a given time. This requires these promoters to have binding sites for multiple TFs allowing for the regulation of a single output with a multi-input parameter, allowing for more complex logic principles to be introduced. With this in mind, we designed two sets of the synthetic promoters with hybrid *cis*-elements composed of binding motifs for MCM1, Mata1, and Matα2 **(Figure 12a)**. Each *trans*-element from these families can interact with the hybrid promoters increasing the number of potential combinations and output range. For pMa_8 specifically, the background expression can be repressed with M2-D-SRDX, while MC-D1-C1 and M1-VP16 increase promoter output at varying levels. This demonstrates expression strength modulation of a single promoter by altering the *trans*-element used to drive its expression. Although the development of our synthetic promoters was intended for protein accumulation, we used this hybrid promoter system to examine the correlation between protein and transcript abundance at two days and five days post-transformation. These data demonstrate a strong correlation (Spearman 0.71) confirming our method of quantifying output at the protein level is sound for observing changes at the transcriptional level **(Figure 12b, 12c)**.

Potentially, the most powerful application of these hybrid promoters is their capacity to introduce multi-gated logic principles into genetic circuits. A simple ‘or’ gate can be generated by utilizing two activating *trans*-elements from MCM1, Mata1, or Matα2, with the promoter in the on state when either activator is present. Additionally, it is possible to develop a reciprocal ‘nor’ gate by utilizing two repressor constructs in concert with a synthetic promoter with high levels of background expression. The promoter remains in the off state when either repressor is bound, only activating when both are absent. Another interesting application of our repressor constructs is the potential to generate a genetic kill-switch that will shut down promoter activity even in the presence of an activating *trans*-element, as demonstrated by the combination of M2-D-SRDX and MC-D1-C1 with pMa_8 **(Figure 12a)**. Utilizing our multi-binding site promoter system, we lay the foundation for the fabrication of more complex and elegant genetic circuits in plants.

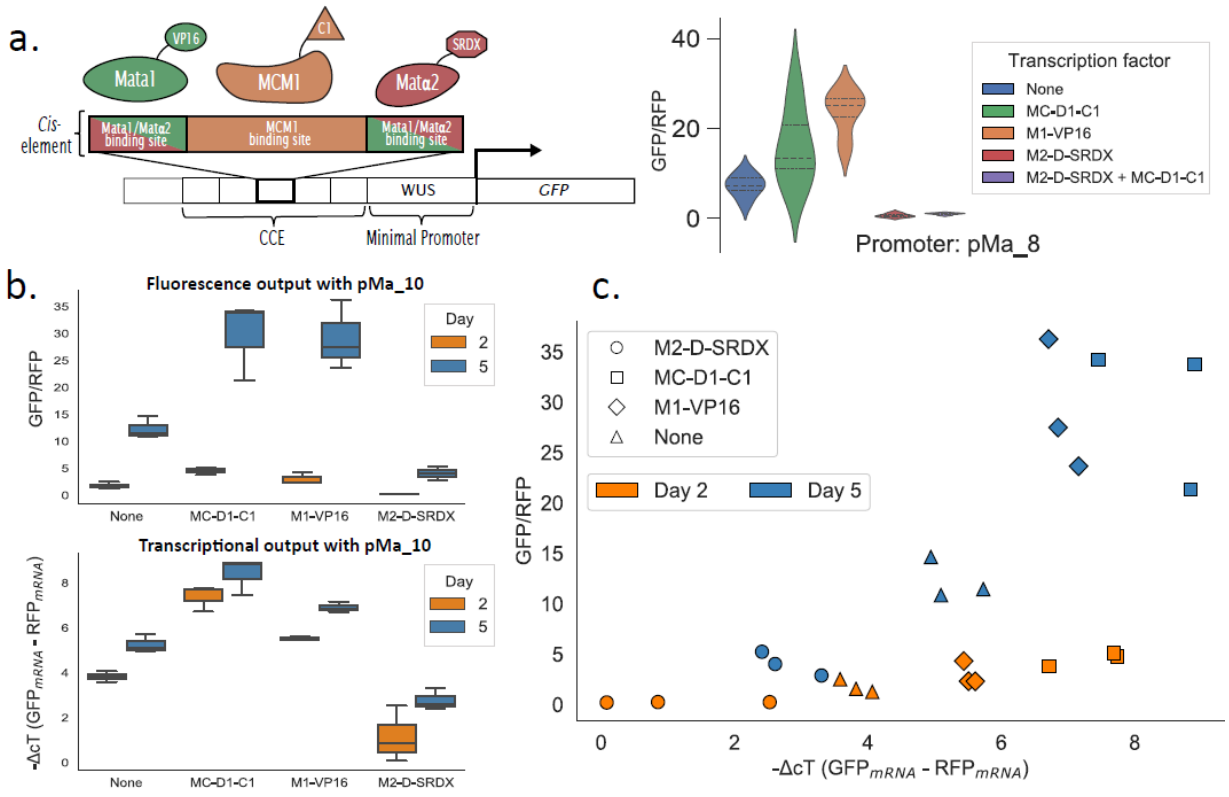


Figure 12. Hybrid promoters incorporate *cis*-elements with binding sites for multiple TF families. a, Graphical representation of the hybrid promoter schematic with *trans*-elements corresponding to the data shown. Varying the family and functional domain of each *trans*-element modulates the expression of the synthetic promoter to varying levels. Additionally, the hybrid promoter can incorporate more than one TF signal simultaneously with output determined by the combination bound. This is demonstrated with the repression of pMa_8 when M2-D-SRDX and MC-D1-C1 are co-expressed with the synthetic promoter. Standard violin plot represents the estimated kernel density with the median and interquartile range represented by dashed lines ($n \geq 8$ biological replicates). **b,** Using hybrid promoter pMa_10, protein and transcript abundance was measured for plants expressing the representative synthetic elements at two and five days. Protein levels were measured using levels of GFP/RFP fluorescence as before. Transcript abundance was determined using qPCR with tissue from the same plants used for fluorescence measurements. The center line of the box and whisker plot denotes the median value, the box represents the interquartile range while whiskers represent the 5^{th} and 95^{th} percentile with values outside considered outliers ($n \geq 6$). **c,** The correlation between protein and transcript abundance was examined with all data from both days. These data have a Spearman correlation of 0.71 with a p-value of 2.09e-8.

Discussion

A major challenge to circuit engineering has been the unravelling and elucidation of genetic determinants underpinning gene expression in eukaryotes, especially in plants. A core tenet of synthetic biology is the ability to understand the fundamental and reductionist rules that govern natural systems in order to reconstruct and engineer artificial molecular components of life. Our findings demonstrate a strategy to investigate contributions of various *cis*-elements, minimal promoters, and *trans*-elements that dictate gene expression, providing the foundation for future studies to rationally design transcriptional regulation systems with predicted expression strengths *a priori*. Although there are many nuances to transcriptional regulation in plants that have not been elucidated in this study, our results take one step towards the coarse dissection of the contributing effect of specific genetic elements in controlling gene expression. These findings may provide the foundation for the future identification of design principles that will enable the construction of more refined and targeted transgene expression systems in plants. It is important to note that the orthogonality of our system may vary in different plant species as imported *cis*-elements may already be present in their genome; however prior work with the Gal4 system in *Arabidopsis* has suggested that there are no major pleiotropic effects from the introduction of new synthetic TFs[124-126]. This will have to be confirmed on a case-by-case basis in future studies utilizing these parts in diverse plant species. Nonetheless, there is potential to expand our conceptual approach for promoter and TF design to other eukaryotes by fusing native core promoters and orthogonal *cis*-elements to build up DNA parts in less developed systems.

A major obstacle that has often thwarted plant engineering efforts is the natural phenomenon of transgene silencing. Plants have evolved robust defense mechanisms that may perceive multiple transgenes driven by the same promoter as a threat, resulting in gene silencing at the transcriptional (gene inactivation via DNA methylation) and post-transcriptional (RNA degradation) level[156,157]. Thus, although many engineering efforts require the coordinated expression of multiple genes, it has long been observed that stacking the same promoter multiple times may also dramatically increase the chance of gene silencing[158]. Inherent to our library design of synthetic elements is the avoidance of identical sequences, as addressed by the randomization of varying CCE combinations and minimal promoters to avoid the use of homologous sequences. This strategy permits the stacking of multiple genes with distinct promoters and may circumvent potential gene silencing. Additionally, it is important to note that many classically used promoters have been incorrectly labeled as 'constitutive,' as they are not expressed in all tissues, nor do they express at similar levels across various tissue types[159]. Thus, this broad stroked approach and traditional reliance on 'constitutive' promoters represents a major barrier that separates the level of engineering complexity that can be deployed in model microbial versus plant systems. A key hurdle has been how to artificially design promoter elements that replicate elements of spatial and temporal control over gene expression[160]. In comparison to microbial efforts, these challenges are unique and specific to plants and have not yet been thoroughly addressed. The majority of previous studies that have required tissue-specific expression of transgenes have utilized endogenous promoters; however, these promoters cannot modulate expression strength, thus demonstrating the one-dimensional limitation of being restricted to these endogenous sequences. The distinct output of each combination of synthetic promoter and *trans*-element enables the expression of numerous transgenes at varying strengths, providing an additional dimension of control over synthetic genetic circuits. Eventually, these

techniques can be expanded to design and build complex multi-layered gene expression systems with synthetic elements

A rapidly growing global population has led to an increase in the motivation and enthusiasm for the engineering of plants to tackle impending societal challenges. Many of these efforts have focused on agronomy, with the goal of increasing crop output per hectare by optimizing plant traits associated with abiotic/biotic stress, disease resistance, biofortification, and sustainability. Additionally, there is a growing interest in using plants for 'molecular farming', with engineered lines grown for the harvest of high-value compounds and proteins. In order to deliver on such agricultural biotechnology solutions, tools will need to be developed that address basic challenges in controlling transgene expression strength, enabling tissue-specific expression, and stacking multiple genes without risks of gene silencing. With these concerns in mind, we have developed a strategy for designing and testing synthetic transcriptional regulators and have demonstrated their utility *in planta*. Expanding the capacity for synthetic plant biologists to perform targeted and precise engineering will be indispensable for future complex multi-gene engineering efforts.

Initial design and testing of synthetic bidirectional promoters

Promoters (in general) are thought to be unidirectional in driving gene expression, often denoted as a single headed arrow, though they inherently show bidirectional behavior in most cases[161]. Generally, the product of upstream transcription is short lived uaRNAs (upstream antisense RNAs), and while currently unknown, may themselves have a function much like siRNAs and miRNAs[162]. While we are not interested in the uaRNAs specifically, this phenomenon demonstrates the capacity for transcription at both the 5' and 3' end of the promoter. This leads us to believe there is potential to engineer a single promoter that drives productive transcription of both the sense and antisense strands of the flanking DNA. This could be quite beneficial when designing synthetic gene circuits with numerous gene-coding sequences by limiting the number of elements required for desired functionality. Especially, when generating GMO plants that require years of work between generations when producing modified lines; and are limited in the amount of transgenic material that can be introduced with each cycle[163]. We theorize that the system we developed for engineering synthetic promoters can be applied for the development of synthetic bidirectional promoters that generate independent mRNA sequences at both ends. In tandem with bidirectional promoter activity, it would be possible to link multiple CDSs with 2A peptide regions to develop polycistronic expression systems for use in plants[164]. This would be especially promising, allowing for the expression of six genes with a single promoter and two terminators. This has the possibility to revise the current approach to plant engineering, which requires a promoter/terminator pair for each CDS. Decreasing the "genetic real estate" used for non-coding sequences like promoters and terminators would increase the capacity for the inclusion of functional CDSs in a single transformation event, accommodating more complex, layered, and stacked engineering schemes.

Bidirectional promoter theory and design

The approach of using DNA-binding sequences (*cis*-elements to form CCEs) and minimal promoter elements (each including a core promoter and a 5'UTR) to generate synthetic promoters was augmented to instigate bidirectional functionality. By appending a second minimal promoter in the antisense orientation upstream from the CCE we aimed to drive productive transcription (resulting in mRNA production of a desired sequence) of both the sense and antisense strand. CCEs of varying length we generated by random concatenation of 1-8 varying *cis*-elements, with randomized 4bp spacer sequences, that bind the Gal4-based *trans*-elements. In place of the minimal promoters used in the initial study, which were sequences derived from various native plant promoters with their corresponding 5'UTR, we selected core promoter sequences developed by *Jores et al.* due to the array of sequence diversity and robust output in their experimental system[165]. Core promoters AT2G39670 and AT3G26090 were selected, as they were top performing elements in their studies using *Agrobacterium* infiltrated tobacco plants. In line with our goal of increasing sequence diversity in our promoter designs, two different 5'UTRs were selected to complete the bidirectional promoter constructs. The *Zm00001d041672* and *HSP18.2* 5'UTRs were selected to flank the core promoters **(Figure 13a)**. The inclusion of the 5'UTR is crucial for the proper processing of the pre-mRNA into mRNA as well as translation. *Zm00001d041672* was chosen due to its inclusion in the *Jores et al.* experiments done in tobacco, (although rationale for its selection was not discussed in their publication). *HSP18.2* was chosen based on previous experiments that demonstrated this 5'UTR promotes robust production and accumulation of both mRNA and protein products[166].

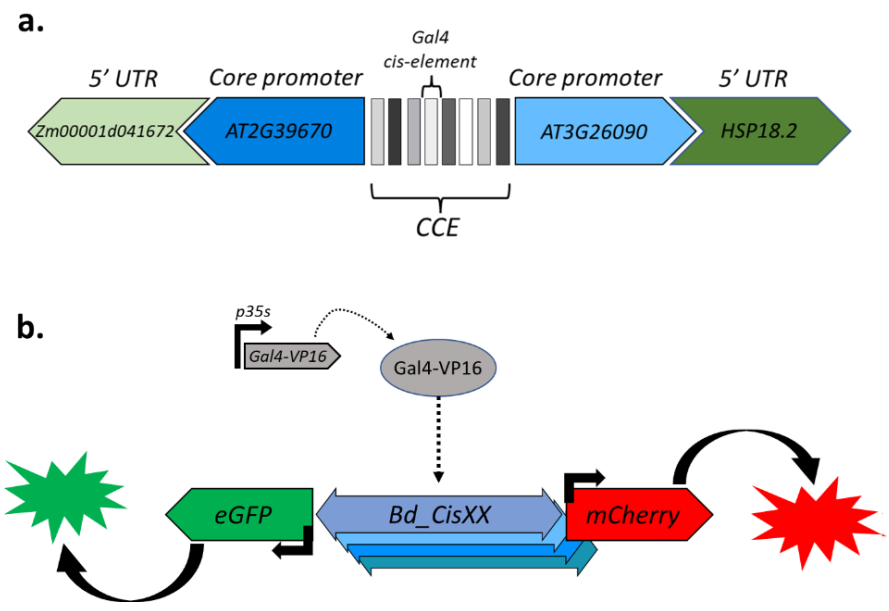


Figure 13. Bidirectional promoter design and testing system. **a,** each individual element of the bidirectional promoter (*cis*-elements, core promoters, and 5'UTRs) has the potential to be modified and optimized to fulfill the needs of the specific application. **b,** Bidirectional promoters were tested by driving expression of eGFP and mCherry, with fluorescent output measured as a proxy for promoter strength. A second binary vector containing *p35s-Gal4-VP16* was used as a transactivation TF to drive expression of the synthetic bidirectional promoter when co-infiltrated into tobacco.

The Gal4-binding *cis*-elements are unique in they are quite palindromic. This feature of the sequence lends to the idea that the zinc-finger domain of Gal4 can bind the DNA in both 5' to 3' and 3'- 5' orientations[167], which makes it a great option for developing a system that functions bidirectionally. The flanking core promoters each bind RNA polymerase and the PIC prior to transcription initiation[168]. Transcription is then activated by the introduction and binding of the Gal4 *trans*-element at the CCE leading to mRNA production at both ends. An additional benefit of this scheme is the almost infinite array of promoters that can be designed by altering the *cis*-elements of the CCE, core promoters, and 5'UTRs **(Figure 13a)**. Each element has an individual contribution to the overall functionality, providing the capacity to tune the strength of expression at each end independently. A two-reporter assay was used to determine functionality of the bidirectional promoters by measuring eGFP and mCherry fluorescence **(Figure 13b)**. Protein accumulation served as a proxy for analyzing the strength of gene expression, this method was verified to correlate with mRNA accumulation in our previous study using these reporters.

Achieving productive bidirectional transcription

While the suite of promoters initially designed and tested did not produce robust results, we were successful in achieving bidirectional transcription of both reporters simultaneously. 37 promoters with CCEs containing 1-8 randomized *cis*-elements were initially tested, showing a range of transcriptional output. All trials were done with the Gal4-VP16 transactivation-element, expressed under the *35s* promoter, in conjunction with an individual bidirectional promoter (denoted as Bd_CisXX). Surprisingly, the number of *cis*-elements used in each promoter lacked significant correlation with output, R^2 value of 0.076 for eGFP and 0.065 for mCherry **(Figure 14)**. Though, further investigation into the individual combinations used in each CCE could glean insight into the contribution of each *cis*-element independently, as there is a subset of promoters that outperform the average. The 5'UTR did show a difference in overall output when comparing the fold-increase in expression over the empty vector control. eGFP, driven by the *Zm00001d041672* 5'UTR end of the promoter, had an average fold increase of 3.31 across all tested promoters **(Figure 14a)**. mCherry, driven by the *HSP18.2* 5'UTR end of the promoter, had an average fold increase of 10.66 across all tested promoters **(Figure 14b)**. *HSP18.2* was shown previously to promote robust transcription/translation in tobacco, likely making it a contributing factor to these observations. The *Zm00001d041672* 5'UTR used on the other end comes from maize. Often, elements from monocots like maize do not function as expected when used in dicots like tobacco[169], which may be an explanation for the minimal output. The core promoters selected to flank the CCE in these designs were shown to have similar strengths in the data set used for selection, making them an unlikely explanation for the major difference in output seen at each end. Though, this does not exclude the possibility that the core promoters themselves were a factor in the differences observed. Overall, the 37 promoters tested did not perform at an optimum level when compared to strong constitutive promoters like *35s*, which is often used to drive high levels of expression in plant systems[170]. While this was disappointing, we do have a framework for optimization of the system via core promoter and 5' UTR modifications.

Insights from initial designs

Even though these first designs did not produce robust levels of output, we were successful in demonstrating bidirectional expression highlighting the potential for engineering bidirectional promoters. We have targeted the core promoters as the likely source for optimization of these primary designs, as additional work done with select *Jores et al.* core promoters found them to

be underperforming in our tobacco system (Dr. Andy Zhou, unpublished). Further comparison of methods also highlighted a key incongruity that is often overlooked. Their measurements were all taken at the mRNA level using barcoded sequences to measure transcription, a sound method for the examination of gene transcription. In our tobacco system, we analyze protein accumulation as a measurement of gene transcription. While both methods are correctly employed, they are measuring different aspects of the central dogma[171]. Inconsistencies are often observed when correlating mRNA abundance and the resulting protein titers. In a theoretical scenario, mRNA with a short half-life and high translation efficiency could produce similar amounts of protein as mRNA with a long half-life and low translation efficiency. If one were to measure the protein abundance, it could be assumed the transcriptional/translational dynamics of both mRNAs were similar, due to similar titers. When measuring the abundance of each mRNA, one would find drastic differences due to varying transcript dynamics and could incorrectly assume this pattern would be observed at the protein level. When selecting core promoters based on mRNA measurements, we inappropriately assumed this would correlate with high levels of protein production, which is the goal for our promoter designs. The second iteration of our bidirectional promoters will shift back to using core promoter sequences mined from dicots, or tobacco specifically, that correlate with high levels of protein accumulation in the native system. These core promoters can then be substituted into our scheme for the optimization of bidirectional promoter design. Additionally, the *Zm00001d041672* 5'UTR would be substituted with varied 5'UTRs from the tobacco *HSP* family to determine the ideal elements for our application.

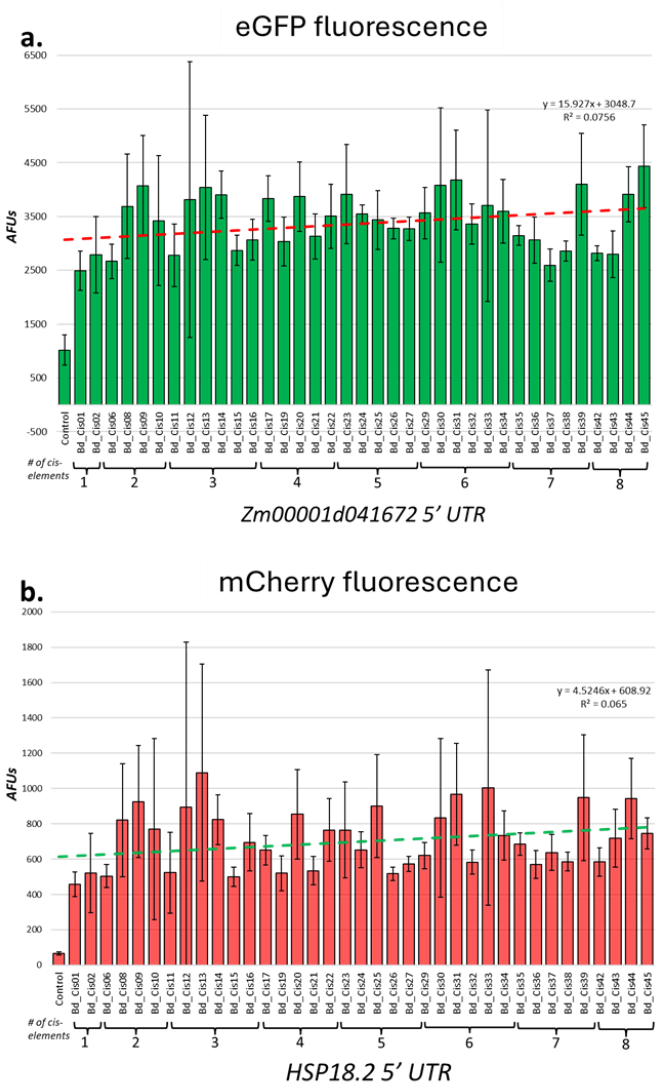


Figure 14. Primary analysis of initial bidirectional promoter designs. a, eGFP flourecence was measured to detmerine output of the 5' end of the bidirectional promoter. Measurements were taken five days post-infiltration (n ≥ 8 biological replicates for each promoter tested, error bars represent the standard deviation from the mean). **b,** mCherry flourecence was measured to detmerine output of the 3' end of the bidirectional promoter. Measurements were taken five days post-infiltration (n ≥ 8 biological replicates for each promoter tested, error bars represent the standard deviation from the mean). The same samples were measured for both eGFP and mCherry expression simultaneously, output was displayed separately.

Increasing the hexose/pentose sugar ratio in sorghum with synthetic biology

As discussed earlier, a major contributor to the recalcitrance of bioenergy feedstocks is the ratio of C6 to C5 sugars. *Sorghum bicolor* is an emerging target of feedstock engineers due its drought resilience and capacity to generate large amounts of biomass[172]. Unfortunately, sorghum feedstocks generally contain high levels of C5 sugar in the form of xylan, which plays a crucial role in maintaining the rigidity of water conducting xylem[173]. Additionally, sorghum has an abundance of xylan-rich sclerenchyma in stem fibers that likely assist in maintaining the rigidity of mature plants[174]. Mutations that disrupt genes responsible for xylan synthesis in model plants often lead to vessel collapse resulting in a dwarfism phenotype[175]. While this would successfully increase the C6/C5 sugar ratio, thus reducing recalcitrance, the decrease in biomass accumulation due to dwarfism would result in a net loss for bioenergy applications. We hypothesize that decreasing the xylan content in the fiber cells specifically, while leaving the xylan of the vessel elements unaltered, would result in an increased C6/C5 ratio of the resulting sorghum feedstock while limiting the dwarfism phenotype. This approach has been successfully demonstrated in *Arabidopsis thaliana* via vessel-specific complementation of xylan biosynthesis in mutant lines[175].

Application of synthetic promoters for the tissue specific engineering of Sorghum bicolor

This work aims to introduce a synthetic genetic circuit into *Sorghum bicolor* that will increase the C6/C5 sugar ratio by decreasing xylan content in the sclerenchyma **(Figure15a)**. Synthetic promoters in conjunction with our tissue-specific expression strategy will be used for the integration of our circuit. A sclerenchyma-specific promoter was mined from the sorghum genome (*pSWN1*) to serve as the integration node. The goal is to have the circuit active only in the sclerenchyma, while silent plantwide, to prevent vessel collapse and therefore dwarfism. *pSWN1*was mined by selecting 2.2kb of the upstream sequence from the *SbSWN1* start codon (LOC8069067), using the NCBI *S. bicolor* genome sequence. SWN1, a homologue of NST1 from *A. thaliana*, is hypothesized to be expressed specifically in the fiber cells, making it a perfect bait to mine a sclerenchyma-specific promoter[176]. This hypothesis is based on research done in the monocot *Zea mays* on the expression pattern and downstream targets of ZmSWN1, demonstrating its role in secondary cell wall synthesis in the fiber cells specifically[177]. The high degree of sequence similarity between SbSWN1 and ZmSWN1 suggests a similar role in both systems, though specific experiments to confirm this in sorghum are yet to be published.

Xylan, a β-(1,4)-linked xylose homopolysaccharide, is biosynthesized by the IRX xylan synthase complex. IRX9 and IRX14 are primarily structural proteins while IRX10 is responsible for the catalytic elongation of the xylan backbone[178]. A dominant-negative suppressor of xylan biosynthesis was generated in *A. thaliana* by mutating IRX10 to abolish the catalytic activity via a G283D amino acid substitution[179]. This caused a drastic dwarfism phenotype when overexpressed plantwide with a *35s* promoter. We designed a similar dominant-negative suppressor for sorghum (IRX10*) by introducing this same mutation at the aligned residue of the sorghum IRX10 homologue (G291D), which is conserved across most land plants. By overexpressing IRX10* in the sclerenchyma we aim to minimize xylan biosynthesis in the fiber cells specifically.

Two different Gal4-based synthetic promoters (*pSbSP1* and *pSbSP2*) were designed using CCE17 and two minimal promoters, one from *Triticum aestivum* and one from *Z. mays*. Combining our top performing CCE in conjunction with previously tested minimal promoters from monocots will provide the highest likelihood of functionality once integrated into the sorghum genome. Unfortunately, we were unable to test these promoters prior to genetic integration of the complete circuit due to a lack of reliable methods for rapid transient transformation and analysis, which is a major challenge in sorghum engineering[180]. The Gal4-VP16 transactivation-element was selected for use in this circuit, having shown robust transcriptional activation in varying experimental designs when coupled with CCE17. The complete circuit will aim to harness the tissue-specific expression pattern of *pSWN1* (serving as the integration node for the circuit) to drive the expression of Gal4-VP16, which is turn binds and activates downstream of *pSbSP1* or *pSbSP2*.

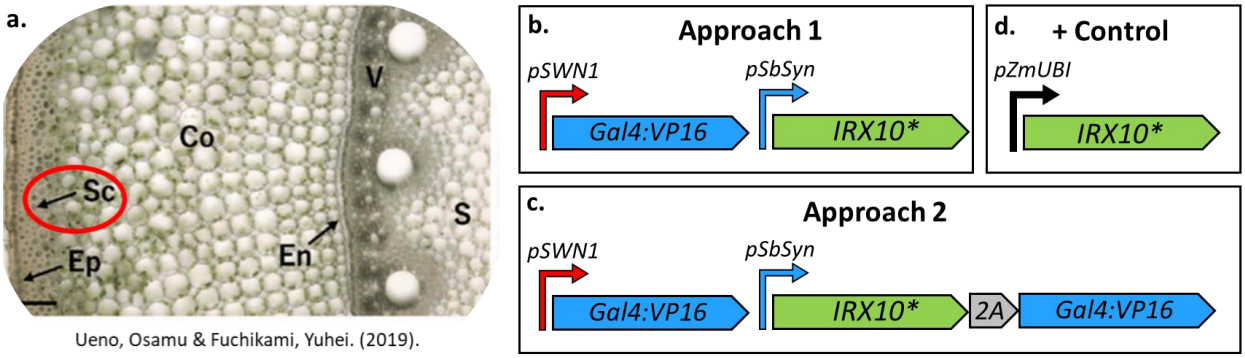


Figure 15. Synthetic genetic circuit schematics for Sorghum engineering.
a, Stem cross-section denoting the location of sclerenchyma. **b,** Approach 1 utilizes a two-component expression system for IRX10* expression in the sclerenchyma. **c,** Approach 2 utilizes a two-component system in conjunction with 2A peptides for bicistronic expression. The aim is to generate a positive feedback loop for IRX10* expression in the sclerenchyma. **d,** Positive control line will express IRX10* plant wide using the constitutive monocot promoter *pZmUBI*.

We designed four iterations of our circuit for the generation of stably transformed sorghum lines for downstream analysis. Two approaches were employed, approach 1 being a simple two-component system, with approach 2 including a positive feedback mechanism. Both synthetic promoters were included with each approach. The two-component approach uses *pSWN1* to drive expression of Gal4-VP16 in the sclerenchyma. Gal4-VP16 then interacts with the synthetic promoter to drive the expression of IRX10* **(Figure 15b)**. With this method the expression of IRX10* is limited by the strength of both *pSWN1* and the synthetic promoter. This could be problematic as IRX10* must outcompete the native IRX10 in complex to disrupt xylose transferase activity. To preemptively hedge against this situation, we employed the second approach to include a positive feedback mechanism using 2A peptides for bicistronic expression[164]. As with approach 1, *pSWN1* drives expression of Gal4-VP16, which interacts with the synthetic promoter to drive expression of IRX10*. To produce the positive feedback loop we included a second copy of *Gal4-VP16* downstream from *IRX10** linked with a 2A peptide sequence **(Figure 15c)**. This second copy of Gal4-VP16 will generate positive feedback by

interacting with the synthetic promoter driving its own expression along with IRX10* to reach the maximum theoretical output. A positive control line was also designed with the aim of inducing the dwarfism phenotype by overexpressing IRX10* plantwide. We aim to achieve this by driving the expression of IRX10* with the constitutive monocot promoter *pZmUBI*, which should lead to the expression of IRX10* in all tissues[*12] **(Figure 15d)**.

Next steps and future experiments

We have outsourced the genetic transformation of our stable sorghum lines to Dr. Ray Collier and his team at the University of Wisconsin. This transformation process takes over a year to generate T_1 sorghum seeds, so unfortunately, we are currently unable to perform an analysis of the physiological effects of our synthetic genetic circuit. Once available, we will germinate and screen for the homozygous insertion of the complete genetic circuit in all five lines. Seedlings that are confirmed for successful transformation can then be grown to maturity and undergo a panel of experiments to determine key phenotypic changes. Specifically, characteristics that have a direct impact on optimizing the feedstock-to-biofuel conversion efficiency will be explored. Various aspects of growth and development will be catalogued throughout the life of the plants including height, biomass accumulation, stem thickness and rigidity, and overall health. Mature plants will then be harvested, and the tissue subjected to numerous chemical and molecular analyses. A monosaccharide analysis will determine changes to xylose levels in specific tissues (stem, leaves, roots, etc.) and plant wide[181]. Additionally, lignin will be quantified and analyzed to examine peripheral changes in secondary cell wall components[*182]. The resulting biomass can then be used as a feedstock for downstream analysis of saccharification and conversion efficiency, which is the ultimate application of the engineered lines. We expect a range of phenotypes in these experiments across all five lines (as well as within each line), with the positive control having the largest increase to the hexose/pentose sugar ratio and the most dramatic dwarfism. The four lines developed with approach 1 and approach 2 will hopefully yield a spectrum of results across all experiments. This will provide a detailed comparison of multiple traits of interest, allowing for a multivariant analysis to determine the most suitable line for bioenergy applications. Eventually, additional traits developed with synthetic biology (i.e., low lignin, decreased cell wall acetylation, and biosynthetic pathways for high value co-products) can be iteratively stacked in the same line to develop an optimum bioenergy sorghum line for implementation in future field trials and large-scale conversion experiments.

Applications of synthetic yeast biology for biotechnology and the biosynthesis of plant natural products

The interest in plant natural products spans the entirety of human civilization, with references of plant-based medical remedies being documented throughout history[183]. Plants have diversified over time to create an array of biosynthetic pathways that produce specialized metabolites with an immense range of chemical structure and function[184]. This can be attributed to their sedentary nature, which drives the evolution of numerous traits that chemically modulate interactions with an ever-changing environment. Advances in genome sequencing/mining has led to the characterization of numerous biosynthetic pathways responsible for the synthesis of alkaloids, terpenoids, phenolics, polyketides, and many more[185-187]. Unfortunately, the large-scale production of many high-value molecules (especially those with pharmaceutical or industrial applications) is hindered in plants. This is mainly due to the considerable energy and time input required to amass a purified product, which is often synthesized at very low concentrations in the native plant host.

Synthetic biology offers a remedy to the challenge of mass production via microbial engineering. Introducing the genes of a biosynthetic pathway for a target molecule into a microbial host delivers a system for the mass production of plant natural products via fermentation in bioreactors. *Saccharomyces cerevisiae* has emerged as an appealing host for the bioproduction of plant natural products due to its well-characterized genome, the understanding of native biosynthetic pathways, the plethora of genetic tools available, and the presence of an endoplasmic reticulum (ER) as a eukaryotic microorganism[188]. The latter is of great importance when dealing with plant natural products specifically, as many of the plant-derived enzymes required for biosynthesis are ER bound (especially cytochrome P450s oxygenases) or interact directly with components of the ER[189]. By optimizing biosynthetic pathway functionality in yeast, it is possible to achieve production titers of high-value plant natural products at far greater a concentration then the native plant host without the need for vast plots of land or fertilizer input.

Pivotal advances in microbial engineering for the biosynthesis of plant-derived terpenoids

Terpenoids are a vast and diverse class of molecules with industrial and medicinal importance. The majority of these molecules are produced across kingdom Plantae via specialized metabolism. Microorganisms, mainly *Escherichia coli* and *Saccharomyces cerevisiae*, have become choice platforms for the biosynthesis of terpenoids due to recent advances in synthetic biology and metabolic engineering. New techniques for gene discovery have expanded our search space for novel terpene synthesis pathways and unlocked unrealized potential for the microbial production of more complex derivatives. Additionally, numerous advances in host and pathway engineering have allowed for the production of terpenoids requiring oxidation and glycosylation, effectively expanding the potential target space. These advances will lay the foundation for the microbial biosynthesis of a seemingly infinite domain of terpenoids with varying applications.

Over 50,000 known terpenoids are naturally produced across all kingdoms of life and constitute the largest class of natural products. Though, the majority are the result of specialized metabolism in plants[190]. Terpenoids are of major interest due to their applications in the food, pharmaceutical, and cosmetic industries as well as their potential as liquid fuels. Microbial biosynthesis is an optimal platform for the isolation and purification of individual terpenoid compounds. Additionally, this limits the need to grow, harvest and extract plant material, which generally contains low concentrations of the desired product along with complex mixtures of similar compounds and prevents the overharvesting of ecologically sensitive species. Terpenes are synthesized from C5 isoprene building blocks derived from the mevalonate (MEV) or 1-deoxy-D-xylulose 5-phosphate (DXP) pathway. These units are then condensed to produce larger and more complex molecules. The major terpene families are the hemiterpenes (C5), monoterpenes (C10), sesquiterpenes (C15), diterpenes (C20), triterpenes (C30), and tetraterpenes (C40). This space is further expanded by the oxidation, via cytochrome P450 oxygenases (CYPs), and glycosylation, via glycosyltransferases (GTs), of the base terpene skeleton/scaffold[191,192]. Furthermore, prenyltransferases add terpene units to other molecules to generate complex terpene composites, or meroterpenoids, which includes the cannabinoids[193] **(Figure 16).**

Within the last two years, large strides have been made in the microbial production of terpenoids from cannabis and hops, with a focus on prenylated chalcones, flavanones, bitter acids, and cannabinoids due to their distinct bioactive properties. All of these products result from the concatenation of a common isoprenoid precursor with a second fatty acid-derived precursor by a prenyltransferase. Prenyltransferases (PTs) in cannabis and hops are severely under characterized yet are responsible for the large diversity within the cannabis & hop terpenomes. A combination of gene network and phylogenetic analyses has been used to identify candidate genes for many of these molecules.

Microbial engineering is currently limited to the biosynthesis of known and well characterized terpenoid products. Computational biology, genomics, and transcriptomics are powerful tools for the identification of novel terpenoid synthesis pathways, ushering in a new era of synthetic biology[194,195]. Yeast has proven to be a desirable host for the production of these complex terpenoids due to its endogenous MEV pathway, the presence of an endoplasmic reticulum (ER) to anchor membrane-associated plant enzymes, and the plethora of genetic tools for the metabolic engineering of this chassis. Bacterial engineering for terpene production has also seen recent success, but there is limited capacity for the expression of CYPs in these hosts without extensive modification of the native enzymes. Here, we focus on recent advances in the discovery of terpenoid synthesis pathways and their heterologous expression in microorganisms, as well as the expansion of the terpenoid target space through host engineering and the utilization of plant-derived CYPs and GTs.

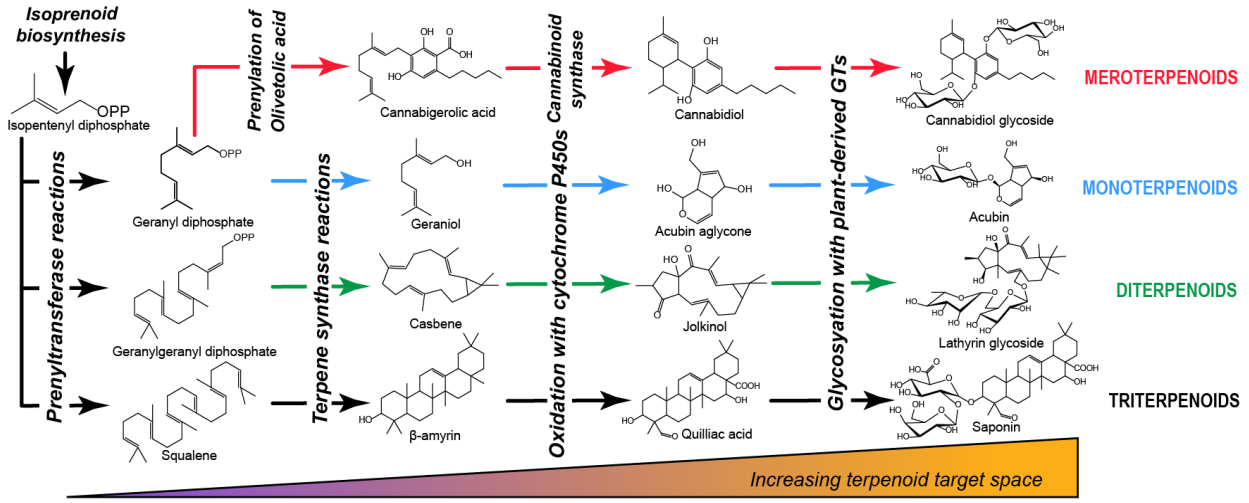


Figure 16. Expanding the terpenoid target space for microbial biosynthesis with plant-derived tailoring enzymes. The molecules used for this representation exhibit a variety of desirable bioactive properties and provide an example of relevant targets for future engineering efforts. There are numerous additional diverse terpenoid families not represented in this figure, and the terpenoids that are displayed represent a single example from potentially thousands in each of the represented families. It is important to note that each arrow does not necessarily signify a single reaction, and in many cases, there are numerous reactions required for the biosynthesis of the representative molecule. Many of the enzymes that catalyze these reactions are yet to be characterized, leaving ample opportunity for future work in pathway discovery and microbial biosynthesis.

Advances in omics studies to identify novel terpenoid biosynthesis pathways

Detailed transcriptomics performed on extractions from glandular trichomes as well as female floral tissues of *Cannabis sativa* at varying developmental stages indicate that candidate genes involved in terpene and cannabinoid synthesis have abundant expression in trichomes[196]. A recent study acquired high quality glandular trichome transcriptomes from nine different commercial cannabis strains in an attempt to characterize terpene synthases[197]. A phylogenetic analysis of the whole genome contig database of *C. sativa* also revealed 24 o-methyltransferases and eight putative aromatic prenyltransferases, one of which was characterized as the penultimate synthase for cannaflavins A and B[198]. Additionally, a total of more than 22,000 expressed sequence tags (ESTs) from several hop trichome-specific cDNA libraries have been deposited in the TrichOME database[199]. This comprehensive transcriptome is a significant resource for further research into natural product pathway discovery and has massively advanced the potential for microbial biosynthesis of multiple terpenoids. Additionally, a recently unveiled metadata platform MaveDB aims to distribute and interpret data from multiplexed assays of variant effect and may be a valuable tool to inform rational engineering[200].

Recent advances in microbial transgene expression systems for terpenoid production

There have been several key advancements in the production of terpenoids from both hops and cannabis in recent years. In 2018, an industrial brewing yeast was engineered to produce primary flavor determinants in hopped beer, allowing for a hoppy tasting beer in the absence of additional hops. Employed was a combinatorial assembly of yeast toolkit parts (promoter, terminator, linker etc.) and iterative design-build-learn-test cycles with strain selection guided by a mathematical model relating genetic design to monoterpene flux[201-203]. To be functionally useful, the engineered strain needed to retain its ability to convert sugars to ethanol, and have precise, stable expression of flavor-determining monoterpenes linalool and geraniol. This work was in contrast to most metabolic engineering efforts which are commonly enlisted to maximize product titers. Multiple state of the art engineering techniques and iterative improvement schemes were employed to tune production of multiple commercially important metabolites without major collateral metabolic changes.

For cannabinoids from *C. sativa*, an aromatic prenyltransferase catalyzes the formation of cannabigerolic acid from olivetolic acid (OA) and geranyl pyrophosphate (GPP). The pathway then branches again toward different cyclized products, such as tetrahydrocannabinolic acid (THCA), cannabidiolic acid (CBDA), and cannabichromenic acid (CBCA). A recent landmark paper describes the successful production of THCA and CBDA from sugar in yeast[205]. Unnatural cannabinoid variants with tailored alkyl chains could also be obtained via feeding the engineered strain with hexanoic acid analogs, demonstrating the substrate promiscuity of olivetolic acid pathway enzymes. Most notably, cannabinoid variants with an alkyne moiety were synthesized, paving the way for future click derivatization. It has been shown that the cannabinoid alkyl side chain is a critical pharmacophore and may be a promising target for pharmaceutical discovery[206]. Another study successfully reconstructed the entire β-bitter acid pathway by heterologous expression of two CoA ligases, a polyketide synthase, and a prenyltransferase complex in an optimized yeast system. A metabolon composed of two aromatic prenyltransferases was elucidated[200]. Another key tool for increasing transgene expression and function for terpenoid biosynthesis is mutagenesis analysis, particularly for prenyltransferases given the plasticity and promiscuity of their active sites[205,207].

Prenylated flavonoids are another subclass of plant phenolics, which combine a flavonoid skeleton with a prenyl side chain. Unlike other flavonoids, they have a narrow distribution in plants, limited to only several plant families, including Cannabaceae. Recent studies have demonstrated that hop terpenophenolics (a term for both bitter acids and prenylchalcones) exhibit diverse bioactivities with a high potential for pharmaceutical applications[208]. A prenylated flavonoid with a very potent phytoestrogen activity is 8-prenylnaringenin, produced in *Humulus lupulus* (hops). 8-Prenylnaringenin was recently produced *de novo* as a proof of concept for yeast as a platform for biosynthesis of prenylated flavonoids[209].

Recently, the importance of non-catalytic foldases and chaperones for terpenoid production in trichomes has been elucidates[200]. THCA and CBDA are unstable and will be non-enzymatically converted to the decarboxylated forms, Δ9-tetrahydrocannabinol and cannabidiol respectively. It is hypothesized that CsaCHIL, a chalcone isomerase-like protein lacking catalytic activity, potentially binds THCA and/or CBDA for stabilization in hemp glandular trichomes and limits negative feedback to upstream enzymes. It has also been shown that upregulation of multiple

foldases and chaperones resulted in a 20-fold improvement of THCA synthase functionality in yeast and poses a promising avenue for optimizing microbial production[210].

Optimizing cytochrome P450 oxygenase function in microorganisms

The progression of terpenoid biosynthesis in microorganisms is limited by the dearth of characterized terpene synthases as well as the CYPs and GTs that modify these terpenes. Computational biology has enabled the discovery of new enzymes, as demonstrated by the identification of 55 predicted terpene synthases from *C. sativa*[196]. CYPs, in particular, are hypothesized to be a main driving force of terpenoid diversification in plants through hydroxylation, sequential oxidations of specific positions (generating anchoring points for additional modifications by transferases), as well as catalyzing ring closure and rearrangement reactions that significantly increase terpenoid complexity[211]. Most CYPs react with a distinct carbon on the terpene backbone, reactions that are challenging for synthetic chemistry, making biosynthesis of oxidized terpenoids a preferable option for production[212]. These CYPs are generally localized to the ER of the native host in close proximity to the terpene synthase producing the substrate for the reaction[213]. Often included on the ER are GTs required for the glycosylation of the oxidized terpenoid, forming potential metabolons on the ER membrane.

There are many inherent challenges with transferring into microorganisms CYPs optimized by nature to work in plant systems. This is a major hurdle when working in prokaryotic cell-factories due to their lack of an ER and cytochrome P450 reductases (CPRs) responsible for transferring electrons between the CYPs and electron carriers in eukaryotes. Groups have successfully engineered *E. coli* with functionally reconstructed plant-derived CYPs by generating fusion proteins with membrane anchors suitable for prokaryotic cells along with the co-expression of a CPR[214-217]. A major advantage of working in yeast systems like *S. cervasiae* and *Yarrowia lipolytica* for the production of decorated terpenoids is the endogenous ER system. This has been successfully demonstrated in *S. cerevisiae* engineered to produce oxidized casbenes, a medically important diterpenoid derivative, that required the optimization of six CYPs, achieving titers of over 1 g/L, building upon techniques initially demonstrated in the landmark paper producing artemisinic acid, a plant-derived sesquiterpene, in yeast[218,219].

Expansion of the terpenoid target space with plant-derived glycosyltransferases

The terpenoid target space can be further expanded through the introduction of GTs from plants into microorganisms for the glycosylation of oxidized terpenoids. Beyond adding new functionality, plants natively produce glycosylated volatile or toxic terpenes for long-distance transport as well as storage of “disarmed" molecules[220]. Saponins, modified triterpenoids synthesized through varying oxiditions and glycosylations of a β-amyrin backbone, have garnered recent interest in both the industrial and human health spaces[221]. The biosynthesis of β-amyrin has been achieved in both *E. coli* and *S. cerevisiae*, but the production of its oxidized and glycosylated derivatives has been limited to yeast[222-225]. Recently, Wang et al. achieved 2.25 g/L production of ginsenoside Rh2, an oxidized and glycosylated triterpene generally harvested from *Panax spp.*, by the directed evolution of UGTPg45[226]. This was the highest titer reported to date for an *in vivo* production system.

Advances in cell-free platforms have enabled the interrogation of GT function *in vitro* and was recently deployed for the production of novel cannabinoid glycosides[227]. This method allows for the characterization of GTs that can then be introduced to a production host for large scale biosynthesis. A challenge for future engineering will be the availability of substrate, nucleotide sugars, for glycosylation reactions in heterologous hosts. Limited work has been done in microbes aimed at producing various nucleotide sugars, but the formation, interconversion, and salvage of these substrates has been extensively studied in plants, providing a framework for future microbial engineering efforts[228].

Increasing terpenoid production titers by altering subcellular morphology

A new paradigm of modifying the subcellular morphology of production cells rather than optimizing metabolic flux has successfully increased oxidized terpenoid production titers in yeast. Kim et al. overexpressed *INO2*, an ER size regulation factor, which resulted in an increase in ER biogenesis, ER protein abundance, protein-folding capacity, and cell growth while limiting ER stress response[229]. This resulted in a 71-fold increase in squalene production and an 8-fold increase in the CYP-mediated production of protopanaxadiol compared to control strains. A similar goal was achieved by knocking-out *PAH1*, which generates neutral triglycerides from phosphatidic acid. This strategy also enlarged the ER and boosted production of β-amyrin, medicagenic acid (oxidized derivative), and medicagenic-28-O-glucoside (glycosylated derivative) by eight-, six- and 16-fold, respectively, over the control strain[230]. These strategies will prove to be pivotal advances in terpenoid engineering and may be applied to any yeast chassis engineered for maximizing the biosynthesis of terpenoids derivatives.

Enhancing product accumulation capacity through host engineering

A potential hindrance of terpenoid biosynthesis in microorganisms is the potential for product or intermediate toxicity preventing the accumulation of high levels of a desired molecule. Achieving maximum accumulation will be essential when commercializing next-generation biofuel alternatives like the sesquiterpene bisabolene[231]. Groups have engineered synthetic hydrophobic droplets within the cell that allow for the storage and accumulation of lipophilic compounds like terpenes while circumventing growth or toxicity issues[232,233]. While this work was done in plants, there is potential to transfer these technologies to microorganisms. Lipid engineering in yeast was accomplished through the overproduction of triacylglycerol and a knock-out of *FLD1*, which regulates lipid droplet size, resulting in oversized lipid droplets that accumulate and store lycopene, an acyclic tetraterpene, resulting in record titers of 2.37 g/L[234]. These challenges have brought recent attention to *Yarrowia* as a production host for plant-derived terpenes due to its capacity to accumulate lipophilic compounds and the potential to utilize technology developed for *S. cerevisiae* in this new host[235,236]. A recent pivotal study harnessed peroxisomes to produce squalene at an unprecedented titer through dual cytoplasmic-peroxisomal engineering[237]. This study indicates that peroxisomes can function analogously to trichomes due to their pathway compartmentalization. While there has been little exploration thus far of the capability of yeast peroxisomes to mimic the trichome metabolic environment specifically, they are a promising avenue for the optimization of heterologous production of terpenoids in yeast.

Conclusion

Utilizing microbial biosynthesis to produce economically relevant terpenoids limits the need to grow, harvest, and extract plant material. This provides an environmentally friendly synthesis platform for specialized terpenoids and permits their production at high concentration and purity. Advances in technologies and strategies for the identification and heterologous expression of terpenoid biosynthesis pathways in microorganisms will provide numerous opportunities for future research. While there has been recent success in engineering prokaryotes for terpene production, yeast will prove to be the optimum production host for more complex terpenoid derivatives and should be a cornerstone for future efforts. The progression of metabolic engineering for terpenoid production is only limited by the identification and application of plant-derived terpene synthases, prenyltransferases, CYPs, and GTs for the biosynthesis and decoration of natural terpenoid scaffolds. By implementing techniques previously described there is potential to expand the latent target space beyond the natural/known terpenome, enabling the biosynthesis of synthetic terpenoids. Achieving this goal will require new breakthroughs in host engineering along with optimizing the expression and function of heterologous pathways. Additionally, generating host strains that produce various or specialized nucleotide sugars for glycosylated terpenoids will provide a chassis for the production of terpenoid glycosides, allowing for the microbial biosynthesis of compounds with altered and enhanced bioactive properties[238,239].

Biosynthesis of taxanes in yeast

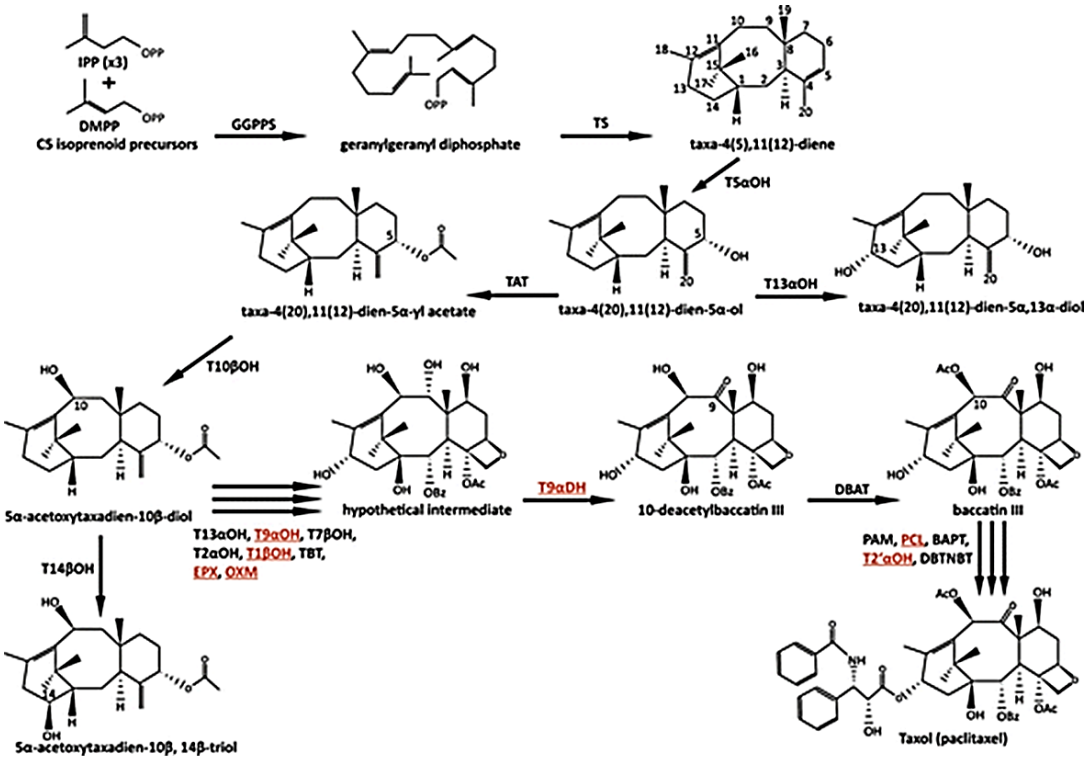
The difficulty sourcing medicinal plant terpenes is exemplified by the Taxol story: clinical development of Taxol was an agonizingly slow progress due to supply shortages of the natural producer *Taxus brevifolia* in the 1980s and 1990s[240]. The concentration of Taxol in the plant is very low (< 0.0001% in bark tissue), and harvesting of yew for extraction is not sustainable, since *T. brevifolia* is now endangered[241]. As is the case for all complex plant terpenes, full chemical synthesis is also not currently a viable economic option as it requires many steps (eleven chiral centers in the case of Taxol), gives low yield, and it not scalable for production. Taxol is currently manufactured either by semisynthesis from 10-deacetylbaccatin III extracted from the needles of *Taxus spp.*, or by extraction from plant cell suspension cultures grown with elicitors to improve production[242]. Both methods still rely on a plant source, resulting in a low and unstable yield, high production costs, and unwanted byproducts. There are many examples of medicinally relevant plant diterpenes that are currently facing similar sourcing issues, with Taxol and cyclopamine as lead examples[243]. This is particularly regrettable because plant terpenes can have unique mechanisms of action not demonstrated by any other class of compounds. For example, Taxol stabilizes microtubules by binding at a unique and specific site resulting in cell cycle arrest making it an effective cancer treatment[244].

There are two major challenges that historically have limited the production of complex plant terpenes in yeast, low yields for the first step in the pathway (terpene synthase reaction) and optimizing complex pathways for the elaboration of the terpene scaffold requiring multiple tailoring enzymes[245]. Previous work with Taxol indicates that multiple products are produced in early stages of the pathway, a major cause of low yields observed in yeast[246]. Additionally, enzymes such as P450s are a notorious challenge for yeast heterologous expression, especially when required to act in series, resulting in diminishing yields of products, thus limiting both pathway discovery efforts as well as the reconstitution of multistep pathways. Despite these challenges, the rational design of strains to tune coupling with redox partners can improve P450 activity in yeast[247]. Along with improving redox dynamics, P450 optimization could be enhanced via augmentation of the ER anchoring regions to improve the localization and expression of plant derived P450s in yeast; or the inclusion of non-enzymatic ER scaffold proteins engineered to bind the P450s for the formation of pseudo-metabolons[248].

Overview of Taxol biosynthetic pathway

Taxol biosynthesis in the native host *T. brevifolia* is a complex pathway requiring nineteen enzymatic conversions, with eight of these enzymes yet to be identified/characterized[249]. This includes eleven ER anchored enzymes (nine P450s, a dehydrogenase, and a cytochrome-P450-reductase) with the remaining predicted as soluble cytosolic enzymes. In *Taxus*, both taxadiene synthase (TXS) and geranylgeranyl diphosphate synthase (GGPPs) are localized to the plastids. This is important to note when introducing TXS in yeast, as localization sequence removal is key to improving performance, with truncation of the first 60aa yielding the largest taxadiene titers[250]. Taxadiene-5α-hydroxylase (T5αH) then converts taxadiene to taxadiene-5α-ol (T5αol), this is the primary modification of terpene scaffold. T5αol is then acetylated at the 5C position replacing the hydroxide. Taxadien-5α-acetate then undergoes numerous oxidations, an epoxidation, dehydration and benzylation to produce 10-deacetylbaccatin III. The final bioconversion of 10-deacetylbaccatin III to Taxol (which can also be achieved via semisynthesis) requires five additional enzymes: phenylalanine aminomutase, β-phenylalanoyl-CoA ligase,

[baccatin III 3-amino, 13-phenylpropanyltransferase], taxane-2α-hydroxylase, and a *N*-benzoyl transferase **(Figure 17)**. Currently, many of the enzymes required for Taxol biosynthesis have yet to be experimentally validated, though multiple research teams are working to characterize the pathway[251]. The daunting task of engineering Taxol production into yeast is made even more challenging due to the spectrum of byproducts generated in the first two steps of the pathway, which drastically diminishes metabolic flux to downstream reactions. The taxadiene to T5αol conversion by T5αH is especially problematic in yeast, producing more than ten dead-end alternative products, with T5αol as a minor product[246]. Prior to this work, the structure of most of these molecules were yet to be experimentally verified which complicates optimization of T5αOH activity and T5αol production.



McElroy, C., & Jennewein, S. (2018). Taxol® biosynthesis and production: from forests to fermenters. In *Biotechnology of Natural Products* (pp. 145-185). Springer, Cham.

Figure 17. Taxol biosynthetic pathway. Taxol biosynthesis from GGPP requires nineteen chemical conversions catalyzed by ER-bound enzymes. As displayed, there are various points in the pathway that can undergo unproductive conversions to dead-end intermediates. The correct "order of operations" for many of the conversions is yet to be experimentally validated.

Deciphering the T5aOH product spectrum in yeast and tobacco

Biosynthetic pathway discovery for plant natural products can be carried out in a variety of systems. Tobacco, specifically *Nicotiana benthamiana*, can serve as an excellent platform for investigating the biosynthetic characteristics of plant derived enzymes with *Agrobacterium*-based transient expression[252,253]. Efforts to engineer plant natural product biosynthetic pathways in yeast can be greatly enhanced when coupled with these tobacco-based investigations.

Especially for high-throughput gene discovery, as plant derived enzymes frequently require modification/truncation for proper functionality in yeast while remaining active in tobacco without modification[254]. We sought to bolster our investigations into T5αH activity with parallel experiments in yeast and tobacco, with the goal of deciphering the T5aH product spectrum and confirming T5aol biosynthesis. Characterizing the array of oxidized products produced by the T5αH reaction with taxadiene is made more perplexing when compared to the GCMS analysis of *Taxus* cells grown in tissue culture. We found that the spectrum of oxidations generated by T5αH is a phenomenon only observed in our transient production systems (yeast and tobacco), with these additional byproducts absent in *Taxus* tissue (Jack Liu, Sattely Lab-Stanford, unpublished). This could suggest a secondary level of regulation in *Taxus* that restricts the "unproductive" oxidation of taxadiene at the 5C position, potentially via allosteric regulation of T5αH, the presence of peripheral (non-enzymatic) proteins that modify reaction kinetics, or substrate shuttling via metabolon formation[255,256]. Interestingly, while the product profile is fairly consistent in both the yeast and tobacco transient systems, the relative and absolute abundance of each molecule are quite diverse.

Taxane production in yeast is carried out in a modified line engineered to overproduce GGPP (JYW001), the primary substrate of the taxadiene diterpene scaffold. TXS was further modified to improve taxadiene production with the removal of the 60aa localization sequence, an N-term maltose binding protein (MBP) fusion, and a C-Term ERG20 fusion (to increase GGPP concentration local to TXS). Two copies of the augmented TXS were integrated into the yeast genome with a galactose inducible promoter to allow for pathway activation after reaching desired cell culture density. After confirmation of taxadiene production, T5αH and a CPR from *Taxus* were integrated into the 2x TXS strain with a galactose inducible promoter for T5αol production. This strain was then used for metabolite extraction and GCMS analysis for comparison to the T5αH product profile generated in the tobacco system, as well as molecular structure confirmation.

More than eleven oxidized products, both mono-oxidized taxadiene (MOTD) and di-oxidized taxadiene (DOTD), are produced in both the yeast and tobacco systems **(Figure 18)**. This demonstrates that not only does the T5αH reaction result in varying oxidations of taxadiene, but also has the capacity for sequential oxidations of taxadiene, further complicating optimization of T5αol production specifically. Unfortunately, the ablation of non-specific T5αH activity will be crucial for the successful integration of the Taxol biosynthetic pathway in yeast. We have currently identified the structure of eight MOTDs and DOTDs produced by T5αH via product purification and NMR analysis. Four of these are MOTDs, one being T5αol and the other three forming epoxides from the hydroxide added at the 5C position **(Figure 18a-d)**. The four DOTDs identified in this work have a myriad of secondary oxidations, with one undergoing a rearrangement of the primary taxadiene scaffold **(Figure 18e-h)**. These data, in conjunction with findings that these alternative products are not found in *Taxus* cells, suggest that the T5αH reaction with taxadiene is not specific in these transient hosts, with the capacity to form numerous transition states during the reaction. This could potentially be attributed to a sub-optimal redox coupling between T5αH and the CPR[257]. Even though both enzymes were mined from *Taxus*, the interaction between them may be altered when outside their native context or orientation. This also strengthens the argument that additional, unknown/unidentified structural proteins or redox intermediaries are present in *Taxus* that modulate the reaction. One difference observed when comparing the GCMS data from tobacco and yeast is the relative abundance of

OCT **(Figure 18b)** compared to T5αol **(Figure 18c).** In yeast, the OCT/T5αol ratio is ~2/1, while in tobacco OCT and T5αol levels are relatively equal (based solely on analysis of peak height). Overall, the differences observed are most likely attributed to altered enzyme behavior (the same polypeptide sequence was used) in yeast and tobacco, as interactions with endogenous proteins and substrates would be quite varied in the plant and fungal host. Comparisons between the absolute abundance and production titers for these products in both systems are not currently viable, as there was no standardization done to normalize between samples. Another important aspect of P450 activity that needs to be explored in future experiments is the balancing of CPR and P450 levels. Overexpression of both the CPR and P450 can lead to a suboptimal redox coupling if the stoichiometry of CPR and P450 are not properly balanced[258]. Experiments that alter the level of CPR expression with constitutive promoters of varying strength (while maintaining overexpression of T5αH) could inform decisions for optimizing the concentration of redox partners, which in turn should improve T5αol production in yeast.

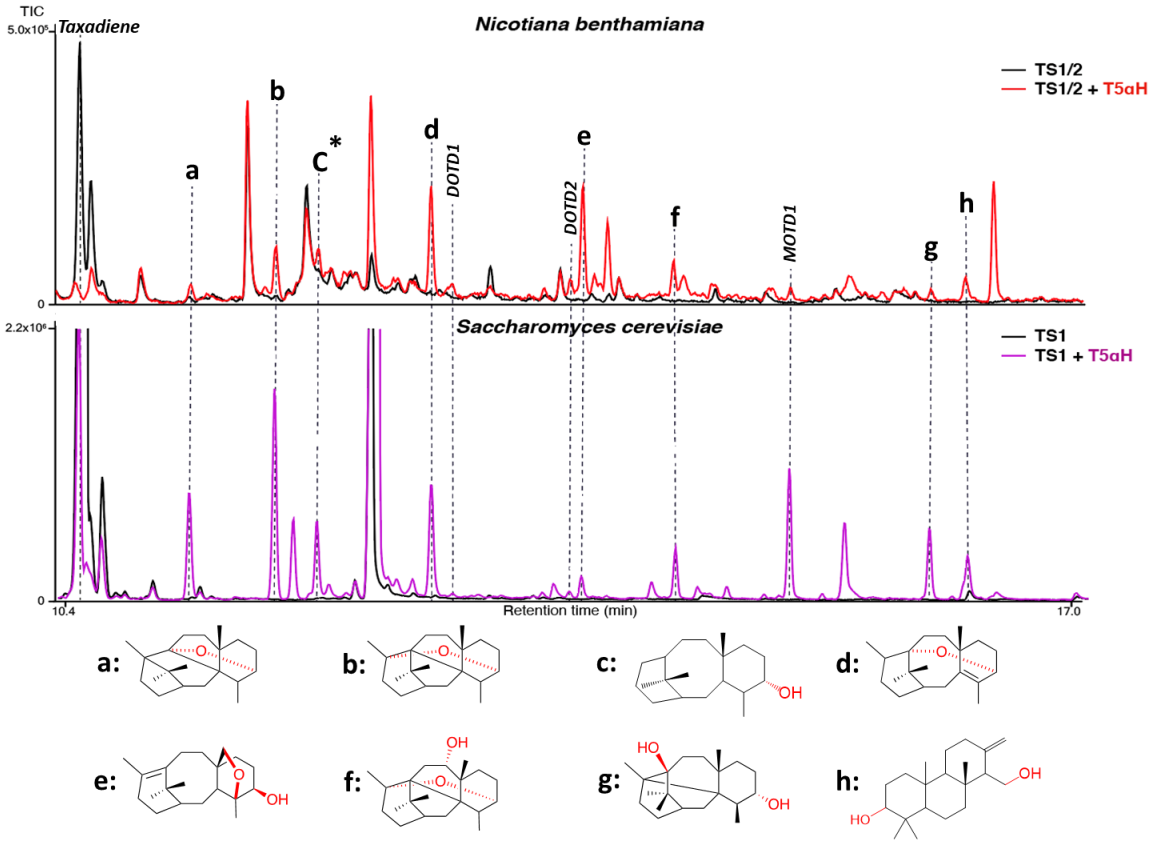


Figure 18. Compilation of T5αH product spectrum in yeast and tobacco. a-d, Four of the MOTDs identified in the GCMS analysis of yeast and tobacco were structurally validated with NMR. **e-f,** Four of the DOTDs identified in the GCMS analysis of yeast and tobacco were structurally validated with NMR. Lines lacking T5αH expression (verified for taxadiene production) are shown in black. The tobacco line with T5αH is shown in red while the yeast strain with T5αH is shown in purple. T5αol (**c***) production was verified via NMR structure analysis as well as comparison to chemically synthesized T5αol (generated with purified taxadiene). Unfortunately, T5αol is a minor product, which demonstrates a major bottleneck in heterologous hosts for the early stages of the Taxol pathway. Additional MOTDs and DOTDs present in the samples are yet to be identified.

Visualization of enzyme expression and localization dynamics

A major challenge when engineering functional P450s from plants into yeast is maintaining robust expression and proper ER localization. P450s, along with their general redox partners CPR and Cytochrome b5 reductase, require ER anchoring for proper assembly and functionality of the electron transport chain coupled to the oxidative reaction carried out by the P450. Confocal microscopy is a great tool for the examination of enzyme expression and localization dynamics when coupled with fluorescent reporter protein fusions[*259]. To better understand how enzymes from plants (specifically from plant natural product pathways) behave in yeast, we generated a suite of fusion proteins consisting of an enzyme of interest and a fluorescent protein for visualization of expression/localization dynamics **(Figure 19)**. In addition to our investigation of specific enzyme behavior in yeast, we sought to better understand how the expression vector used in each experiment (plasmid-based expression vs. chromosomal expression) influences protein production across a population.

A common method for enzyme characterization in yeast utilizes high-copy plasmids with an auxotrophic selection marker for expression[*260]. Unlike most antibiotic-based selection methods, kanamycin selection in *E. coli* for example, auxotrophic selection has a degree of inefficiency. It is possible for "cheaters" not carrying the plasmid containing the selectable marker to persist by scavenging the deficient amino acid from dead cells in culture. When comparing the expression patterns of an integrated fluorescent protein (eYFP) and various fusion proteins carried on an auxotrophic plasmid we observed a drastic effect on expression across the population **(Figure 19)**. Using either of the plasmid-based vectors (URA or LEU auxotrophic selection) for fusion protein expression resulted in a mixed population when visualized, specifically when analyzing the binary pattern of expression vs. no expression in each cell **(Figure 19b-d)**. While other factors likely contribute to this phenomenon, there is a consistent pattern observed when comparing the expression of stably integrated and plasmid-based expression vectors. Even though genetic integration of an expression vector yields homogenous expression in a population, and thus consistency in production titers and peripheral observations, plasmid-based expression has many utilities when engineering biosynthetic pathways in yeast. While not advised for strains engineered for large-scale production, plasmid-based expression is excellent for initial experiments aimed to characterize unknown biosynthetic enzymes. This is due to the high throughput manner in which enzymes can be tested and optimized prior to genetic integration, providing a primary system for investigating novel biosynthetic pathways in yeast while building a final production strain.

A key observation can be made when comparing the localization pattern of a full-length dioxygenase (the complete CDS from the native host including a predicted localization pattern) with that of the same enzyme with the localization sequence truncated. Expression of the full-length sequence in yeast results in a unfunctional enzyme, which is corroborated by the abnormal localization pattern. Truncation of the native localization sequence restores biosynthetic functionality (Dr. Graham Hudson, UC Berkeley-Keasling Lab, unpublished) as well as the proper cytosolic localization in yeast **(Figure 19c,d)**. These data demonstrate the utility of confocal microscopy for the visualization of enzyme behavior, which can inform the interpretation of production titers when rebuilding biosynthetic pathways in yeast.

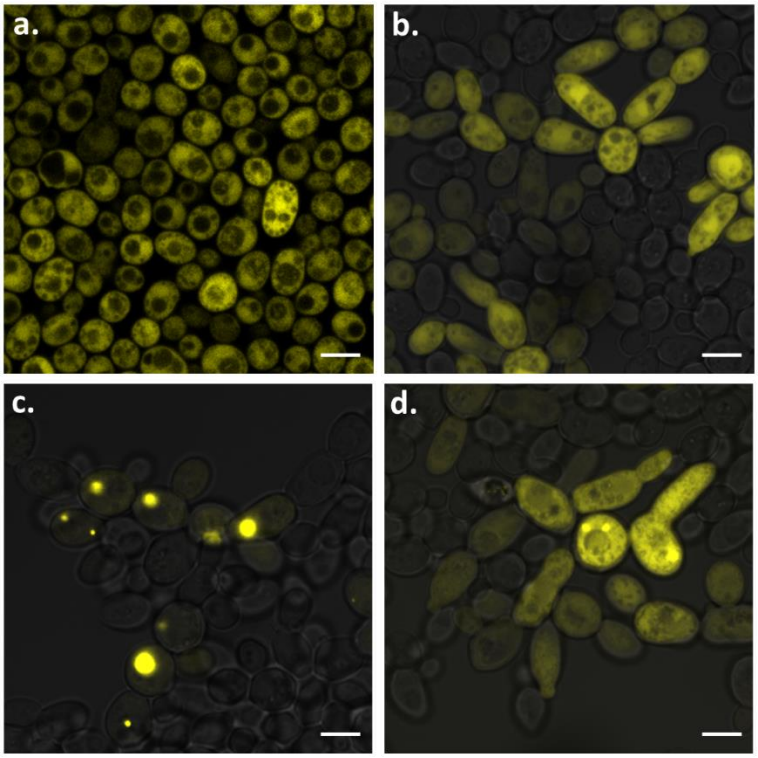


Figure 19. Comparison of expression with genetic integration and auxotrophic plasmid vectors. a, eYFP is expressed via genetic integration, signal is present in all cells. **b,** A cytosolic double bond reductase fused to eYFP is expressed from a URA auxotrophic plasmid. This shows proper cytosolic localization, though expression is limited to <50% of the population. **c,** A cytosolic dioxygenase, full-length native sequence fused to eYFP, is expressed from a LEU auxotrophic plasmid. Protein fails to properly localize and accumulates to unknown cellular bodies; expression is limited to <30% of the population. **d,** The cytosolic dioxygenase with eYFP fusion has been truncated to remove the predicted plant localization sequence. This restores proper cytosolic localization in yeast when expressed from a LEU auxotrophic plasmid, though expression in still limited to <50% of the population. These findings highlight the inefficiency of expression from auxotrophic plasmids as well as the effect of enzyme modification when introduced into yeast. Cultures were visualized 8hr PI at 1000x, scale bars represent 10uM.

Another aspect of enzyme expression and localization we wanted to explore was stability over time. Taxadiene-5α-acetyltransferase (TAT), a soluble and cytosolic enzyme from the Taxol pathway that acetylates T5αol, was fused to eGFP and visualized over time post-induction **(Figure 20a-c)**. At 8hr post-induction (PI) TAT:eGFP is seen properly localized to the cytosol with minor puncta forming in a few cells. Though, when visualized at 24hr PI the majority has been shunted into numerous puncta in each cell, with limited fluorescence remaining in the cytosol. Complete amalgamation of TAT:eGFP into these puncta can be seen at 48hr PI. This could be an explanation for the lack of biosynthetic activity for TAT when introduced into the T5αH production strain, though minimal levels of T5αol (or other unidentified issues) could also restrict activity. While the mechanism responsible for this shift from dispersed cytosolic localization to compartmentalized bodies has not been identified, it is hypothesized that the enzyme is being trafficked into proteosomes or peroxisomes for turnover. Especially when compared to a peroxisome localized fusion protein[261], which shows a similar (though more dramatic) pattern seen at 48hr PI **(Figure 20d)**. Observing this localization pattern over time has highlighted additional aspects for optimization when engineering biosynthetic pathways in yeast, persistence of localization and stability/half-life. Ensuring that the enzymes required for target molecule biosynthesis have robust expression, along with persistence in both localization and half-life, is crucial when optimizing production strains for large scale synthesis schemes.

An additional characteristic of enzyme behavior in yeast was observed when comparing localization dynamics over time of an ER and a cytosolically localized enzyme. Ketoreductase 11 and 23 (KR11 and KR23) involved in saponin biosynthesis (a plant natural product from *Quillaia Saponaria* requiring numerous oxidations and glycosylations[262]) were selected for visual comparison. KR11 (cytosolic localization) was tagged at the C-term with eYFP and KR23 (ER localization) was tagged at the C-term with CFP to analysis localization simultaneously. KR11:eYFP and KR23:CFP were cloned into the same plasmid-based expression vector with Gal1 and Gal10 (galactose inducible promoters) driving their expression. When visualizing the localization of both KR11:eYFP and KR23:CFP at 8hr PI, KR11:eYFP is seen properly localized to the cytosol while KR23:CFP is properly localized to the ER membrane **(Figure 20e)**. After confirming both enzymes were expressed and properly localized at 8hr PI, we visualized the same culture at 24hr PI. As previously observed with other cytosolically localized enzymes, KR11:eYFP had been shunted into small bodies with very little remaining in the cytosol. KR23:CFP on the other hand retains proper ER localization in most cells, while showing limited loss in overall abundance **(Figure 20f)**. This phenomenon demonstrates an incongruity in behavior between ER-anchored and cytosolically localized enzymes from plants when expressed in yeast. The ER localized enzymes are largely occluded from accumulation to the "unknown bodies" that capture the cytosolic enzyme. While the mechanism of action responsible for these observations has yet to be characterized, these findings highlight aspects of enzyme behavior that should be explored when engineering biosynthetic pathways from plants into yeast.

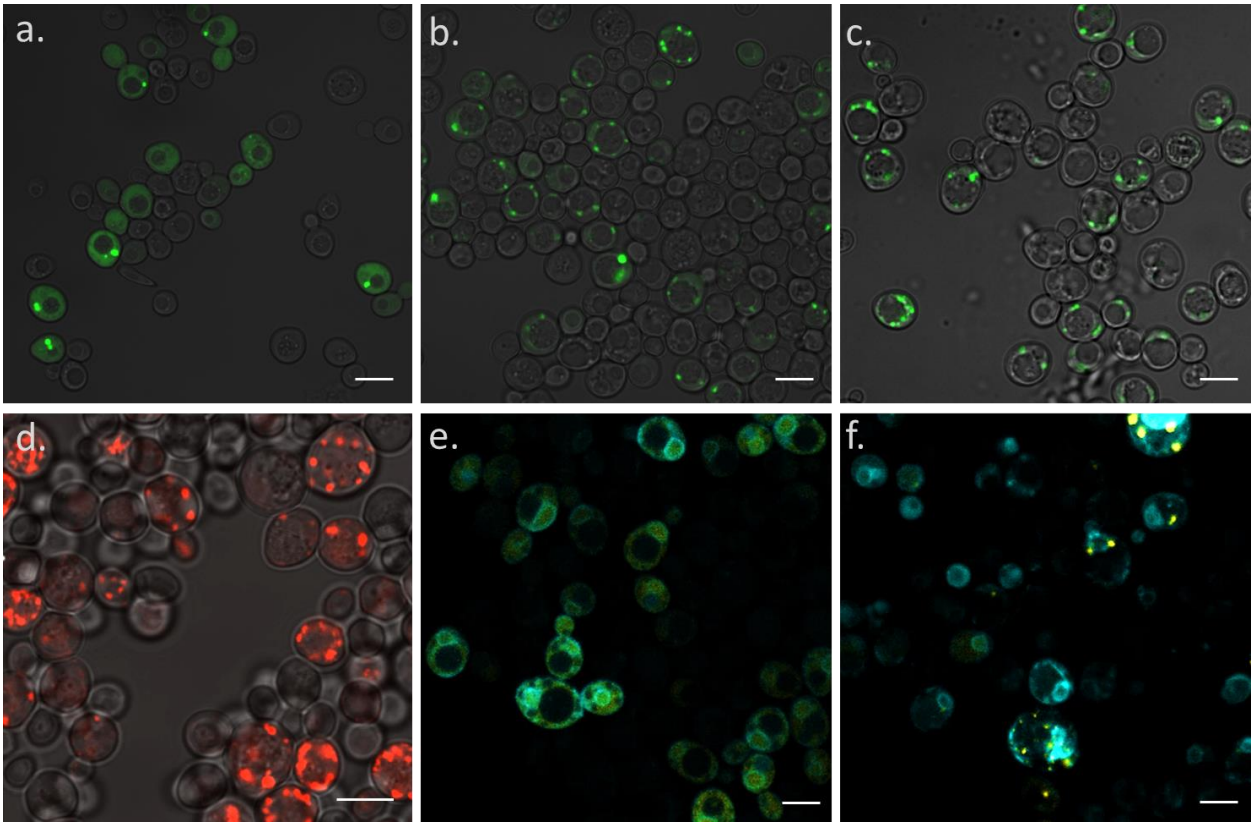


Figure 20. Observations of enzyme expression/localization in yeast over time. **a,** TAT:eGFP expression from an auxotrophic plasmid was visualized at 8hr PI. Cytosolic localization is observed in the cells expressing the fusion protein (retaining the auxotrophic plasmid). The same TAT:eGFP culture was visualized at 24hr PI (**b**) and 48hr PI (**c**). Over time, the cytosolic localization of TAT:eGFP shifts as the protein is accumulated in unidentified cellular bodies, potentially peroxisomes for protein turnover. **d,** GalT:mCherry was tagged with a peroxisome localization sequence to visualize peroxisome localization dynamics of a fusion protein. When compared with **b** and **c**, it can be hypothesized that the TAT:eGFP fusion protein is being shunted to peroxisomes over time via an unidentified mechanism. **e,** KR11:eYFP and KR23:CFP were co-expressed and visualized at 8hr PI. KR11:eYFP properly localizes to the cytosol while KR23:CFP properly localizes to the ER membrane. **f,** The same culture visualized in **e** was sampled at observed at 24hr PI. KR23:CFP maintains proper ER localization (in most cells) while KR11:eYFP can be seen accumulating in the same unknown bodies demonstrating a variation in behavior based on localization. Cultures were visualized at 1000x, scale bars represent 10uM.

Taxol-associated P450 localization in yeast

Unfortunately, the P450s required for Taxol biosynthesis fail to properly localize to the ER membrane when expressed in yeast. This issue presents a major challenge for engineering a functional Taxol biosynthesis pathway due to the required redox coupling with ER bound enzymes. One option for reconstituting proper ER localization is to exchange the native ER anchor region of the P450 with an anchor sequence known to properly bind the ER membrane in yeast[263]. This chimeric protein would hopefully have restored ER localization while retaining the

native biosynthetic characteristics. This is more challenging with a subset of the Taxol P450s that are predicted to have multiple transmembrane domains with an uncharacterized ER-binding mechanism. Modifying this interaction without disrupting the internal loop structures and catalytic activity of the protein could prove quite difficult. Another aspect of P450 function is often determined by peripheral protein-protein interactions in the native host. Many highly modified plant natural products are synthesized by a collection of enzymes in complex along the ER membrane called metabolons, with the intermediate substrates shuttled from one reaction center to the next. These highly optimized systems often involve non-catalytic "scaffold" proteins that act as a foundation for the binding and assembly of numerous enzymes.

Examining the expression and localization dynamics of Taxol-associated P450s revealed a deficiency in ER localization when tested in their native form in yeast. Expression was seen for all the P450s investigated, but they localize to the cytosol. As mentioned earlier, functionality for the vast majority of P450s from plants are dependent on proper ER-anchoring and redox coupling with additional ER-bound enzymes. We sought to modify the localization of Taxol P450s in yeast by substituting the native ER-anchor region (predicted *in silico*) for a sequence previously shown to result in proper ER localization in yeast. Structure modeling and transmembrane domain prediction was carried out for six of the Taxol P450s (T2αH, T5αH, T7βH, T10βH, T13αH, and T14βH) using the PHYRE2 platform. In general, the first transmembrane domain of a P450 acts as the ER-anchor region. We therefore selected the first amino acid after the predicted transmembrane domain for each P450s as a truncation point. The functional ER-anchor sequence (MEHLYLSLVLLFVSSISLSLFFLFYKHKS) was used to replace the truncated region along with the same linker sequence (GSAGSAAGSGEF)[264] used for the fluorescent reporter protein fusions. The linker sequence was added with the aim of improving ER localization by introducing flexibility at the fusion site. Unfortunately, even with the modified anchor region the P450s fail to properly localize to the ER **(Figure 21)**.

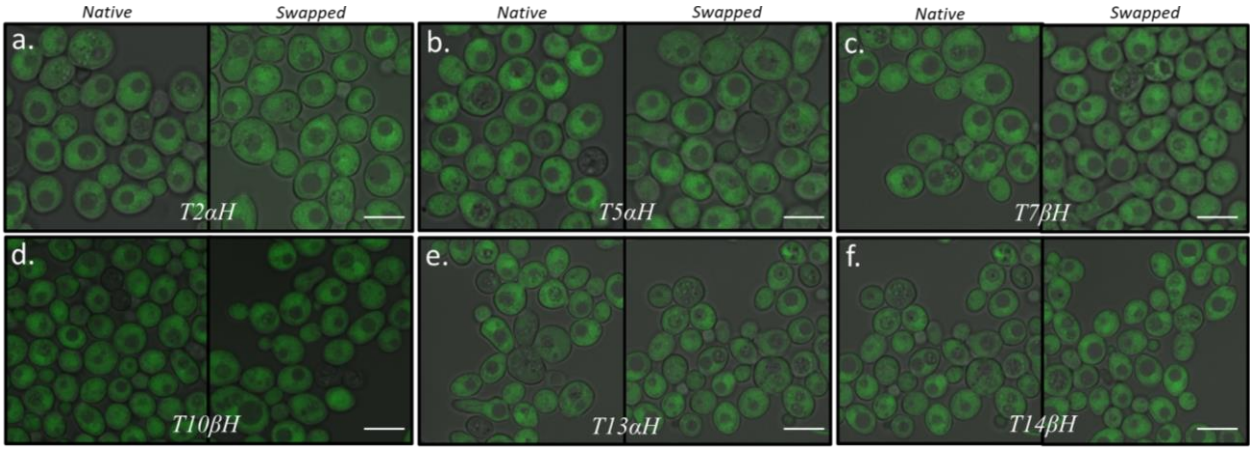


Figure 21. Functional ER-anchor swap fails to induce proper ER localization of Taxol P450s. Six of the P450s required for Taxol biosynthesis were tested for ER localization in yeast. All six P450s failed to properly anchor to the ER membrane, remaining in the cytosol. By swapping the predicted ER anchor of these P450s (first transmembrane domain of the polypeptide) for a functional ER-anchor sequence, we hoped to induce proper localization for these enzymes. Unfortunately, the modified P450s still fail to properly localize to the ER membrane. All P450:eGFP constructs were genetically integrated into the taxadiene production strain for expression and localization visualization. Cultures were visualized 8hr PI at 1000x, scale bars represent 10uM.

Another potentiality for improving the localization and function of multiple P450s is engineering protein scaffolds that promote assembly at the ER membrane. Membrane steroid binding proteins (MSBPs) are a class of protein that are commonly involved with biosynthetic pathways in plants[265]. Furthermore, these are often required for functional biosynthesis, exemplified by the lignin biosynthetic pathway involved in secondary cell wall formation. Biosynthesis of the lignin polymer is highly regulated, multi-branched, and includes several P450s requiring a MSBP for complex assembly[265]. By identifying transmembrane domain anchors and MSBPs with a high affinity for interaction, it may be possible to generate pseudo-metabolons when engineering plant natural product biosynthetic pathways into yeast, specifically those that require numerous ER-bound enzymes. To test the potential of this application, we selected a MSBP to test in yeast with a previously characterized fusion protein confirmed for functional ER localization (ER-CslG). Fluorescent reporters were used to observe the expression and localization of both MSBP (MSBP:mCherry) and ER-CslG (ER-CslG:eGFP) simultaneously to examine co-localization patterns **(Figure 22)**. ER-CslG:eGFP showed proper ER localization patterns in most of the cells observed **(Figure 22a)**, while MSBP:mCherry showed varied patterns of localization **(Figure 22b)**. It is important to note that there are potential unknown interactions between MSBP and the mCherry fusion protein. In the cells that showed high levels of co-localization though, robust accumulation of both MSBP and ER-CslG can be seen along the ER envelope of the nucleus **(Figure 22d)**. The pattern observed seems to represent a nucleation reaction, with co-localization initiation leading to the accumulation of both MSBP and CslG at high concentrations. This observation exemplifies a prospective method to promote the accumulation and optimum localization of plant derived P450s, as well as providing a secondary scaffold for the “assembly”

of multiple P450s simultaneously. While more optimizations would be needed to integrate numerous enzymes into this engineering scheme, the potential of forming pseudo-metabolons could be a favorable approach when engineering biosynthetic pathways of plant natural products like Taxol in yeast.

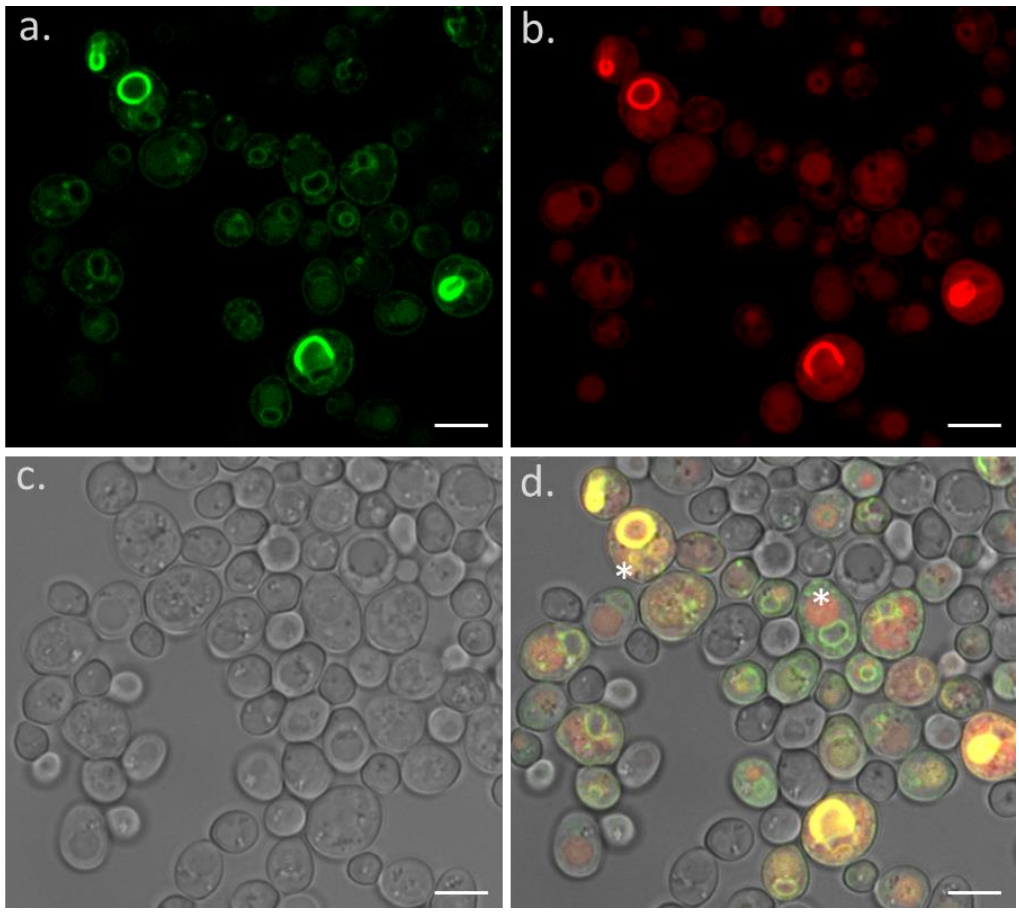


Figure 22. Incorporation of a MSBP induces accumulation and co-localization of ER-anchored P450. a, Visualization of ER-CslG:eGFP expression and localization. **b,** Visualization of MSBP:mCherry expression and localization. **c,** Brightfield image of the cell culture. **d,** Merged image compiling **a**, **b**, and **c** shows a high degree of co-localization of MSBP and ER-CslG. Cells with the highest amount of co-localization demonstrate a robust accumulation of both MSBP and ER-CslG at the ER membrane. Though, there is a mixed population within the culture that fails to colocalize. Individual cells exemplifying each condition are highlighted with a while star. Cultures were visualized 8hr PI at 1000x, scale bars represent 10uM.

Biosynthesis of steroidal alkaloids in yeast

Steroidal alkaloids are a class of molecules with a diverse range of structure and physiological effect[266]. They are commonly found in plants from the nightshade family, genus *Solanum,* including tomato, eggplant, and potato. Yeast provides a relatively unexplored opportunity for the bioproduction of steroidal alkaloids in a heterologous host. As yeast does not natively produce cholesterol, the primary substrate for steroidal alkaloids, very limited efforts have been made to explore this molecular space for bioproduction. Interrupting ergosterol production and shunting zymosterol to cholesterol can be achieved by genetically exchanging *ERG5* and *ERG6* with *DHCR7* and *DHCR24*[267]. Additionally, the substitution of cholesterol for ergosterol in the cell membrane results in limited disruption on yeast growth and development, making this a plausible system to explore. This "humanized" yeast system has been used in studies of mammalian cell-surface receptors and transporters in the past but harnessing this chassis for the bioproduction of plant natural products has untapped economic potential[268]. Further modifications could be made to optimize cholesterol overproduction and bioavailability, to ultimately produce a *S. cerevisiae* strain engineered specifically for the biosynthesis of cholesterol derivatives with diverse structures and applications.

Development and characterization of cholesterol producing yeast chassis

To explore the potential of yeast as a platform for steroidal alkaloids production, we engineered a cholesterol producing strain using CEN.PK as our parent line[269]. *DHCR7* and *DHCR24* from humans were selected based on previous work on cholesterol biosynthesis. Sequential gene substitution of the native *ERG5* and *ERG6* genes was carried out via homologous recombination. *pADH1-DHCR7* was used to replace *ERG5* and its native promoter with *pADH1-DCHR24* replacing *ERG6* and its native promoter, gene substitutions were performed at the native loci. One deficiency of "humanized" yeast is the capacity for acid transport, as cholesterol substitution disrupts endogenous acid transporters[267]. This was observed when performing counter selection for the URA auxotrophic marker used in cloning, which utilizes plating on 5-FOA (final medium pH of ~3.5). This acidic medium, in conjunction with the 5-FOA counter selection (a weak acid requiring the acidic pH to maintain protonation and thus functionality), caused a drastic lag in growth for the final strain with colonies appearing more than ten days after plating. Multiple colonies that cleared PCR and sequencing screens were then selected for GCMS analysis of cholesterol production against an analytical standard. Interestingly, two strains were confirmed for cholesterol production (*Chl*24-20 and *Chl*75-10) but have varied product profiles. *Chl*24-20 was the top producing strain in this initial screen, with cholesterol being one of two major products in the GCMS analysis. While *Chl*75-10 did produce cholesterol, it was a minor product, with two other major peaks present **(Figure 23a)**. These were not structurally identified but based on MS spectrum analysis are hypothesized to be structurally similar to cholesterol with variation in the number of C-C double bonds. *Chl*24-20 was selected as our production strain for downstream analysis.

While final cell densities of experimental strains and the control strain were equivalent at 72 hours, there was an observable lag in growth for the experimental strains during the first 36 hours of production. This is most likely due to sub-optimum cholesterol production along with peripheral effects of sterol substitution in the developing membranes of dividing yeast. After confirmation of cholesterol production in *Chl*24-20 with an analytical standard, we used GCMS analysis of *Chl*24-20 culture extract and a series of five dilutions of cholesterol standard in ethyl

acetate, which is the same final solvent used for extraction **(Figure 23b)**. These data were then used to generate a standard curve for cholesterol to quantify production in *Chl*24-20, which was calculated to be ~128.4nM with our extraction method **(Figure 23c)**.

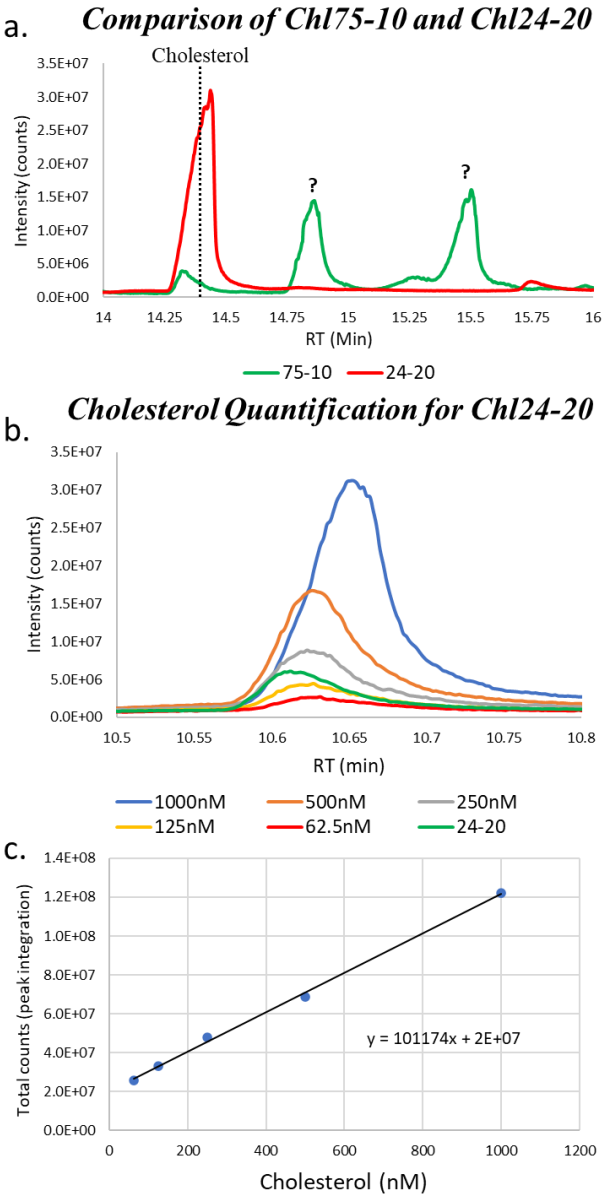


Figure 23. Confirmation and quantification of cholesterol production.

a, Two lines confirmed for cholesterol production were compared via GCMS. *Chl*75-10 produces limited cholesterol in conjunction with two other unknown products. *Chl*24-20 produces more cholesterol without these other products, and thus was selected for further analysis. While not confirmed, based on MS analysis the unknown peaks are hypothesized to be structurally similar to cholesterol with a variable number of C-C double bonds. This observation is most likely due to an incomplete removal of *ERG5* and *ERG6* when developing the strain, leading to the production of these cholesterol-like molecules with varying bond structure. **b,** A serial dilution of a cholesterol analytical standard was used to develop a standard curve for quantification of cholesterol production. **c,** Cholesterol production in *Chl*24-20 was quantified. Based on the integration of peaks shown in **b**, we calculated cholesterol production in *Chl*24-20 to be ~128.4nM with current extraction methods.

Future potential for glycosylated steroidal alkaloid production in yeast

Numerous steroidal alkaloids have implications in human health and research including cyclopamine, α-tomatine, and α-solanine[243,270,271]. The interest in α-tomatine (native to tomato) is due to its function as an antimicrobial that protects the plant against fungi as well as herbivores[270]. Its broad-range mechanism of action, which involves disruption of cellular membranes as well as inhibition of acetylcholinesterase, makes it a great candidate for further development of antibiotics. α-solanine, from potato, has a similar membrane disruption effect, but is different in its capacity to interact with mitochondrial membranes. It acts as a poison by disrupting the mitochondrial membrane potential leading to a flood of calcium ions into the cytoplasm causing cell damage/death[272]. Cyclopamine is of particular interest due to its interaction with Hedgehog signaling, which is involved with an array of developmental processes including embryonic development and tumorigenesis[243]. Much like Taxol, it has profound implications for its capacity to inhibit tumor growth as a treatment for a variety of cancers, though the mechanism of action is quite different. Cyclopamine is a steroidal alkaloid derived from cholesterol that does not require glycosylation, making it a great target to test the potential of our steroidal alkaloid yeast platform. Verazine is an intermediate in cyclopamine production that has a well characterized pathway requiring three P450s and an aminotransferase (*CYP90B27*, *CYP94N1*, *GABAT1*, and *CYP90G1*)[273]. The conversion of verazine to cyclopamine then only requires two additional reactions that are currently being charactered (Dr. James Reed, John Innes Center-Osbourne Lab). What is exciting about this pathway is the potential to generate novel glycosylated steroidal alkaloids with the introduction of GTs from either saponin, tomatine, or solanine biosynthesis for the glycosylation of the 3C position **(Figure 24)**.

Recent advances in engineering nucleotide sugar pathways into yeast, coupled with the promiscuous nature of plant derived GTs, provides an opportunity for the biosynthesis of glycosylated steroidal alkaloids[274,275]. Solanaceous plants produce a variety of these molecules, with diverse physiological applications and effects and therefore have an abundance of genetic material that can be implemented in yeast[276]. The potential to produce novel glycosylated molecules is especially promising, for example glycosylated cyclopamine derivatives, with the introduction of GTs that glycosylate similar structures such as saponins. These new molecules have a greater degree of structural diversity and may provide enhanced, altered, or even novel properties. This specifically demonstrates the utility of synthetic yeast biology as a tool to produce “new-to-nature” molecules not possible with classic organic chemistry or extraction from plants. While further optimizing the production of free cholesterol, we can begin to introduce key nucleotide sugar pathways that produce the substrate for GT reactions, allowing for pathways requiring glycosylation to be investigated. Engineering a yeast platform specifically for the bioproduction of glycosylated steroidal alkaloids would be an excellent resource for future studies on these complex and derivatized molecules

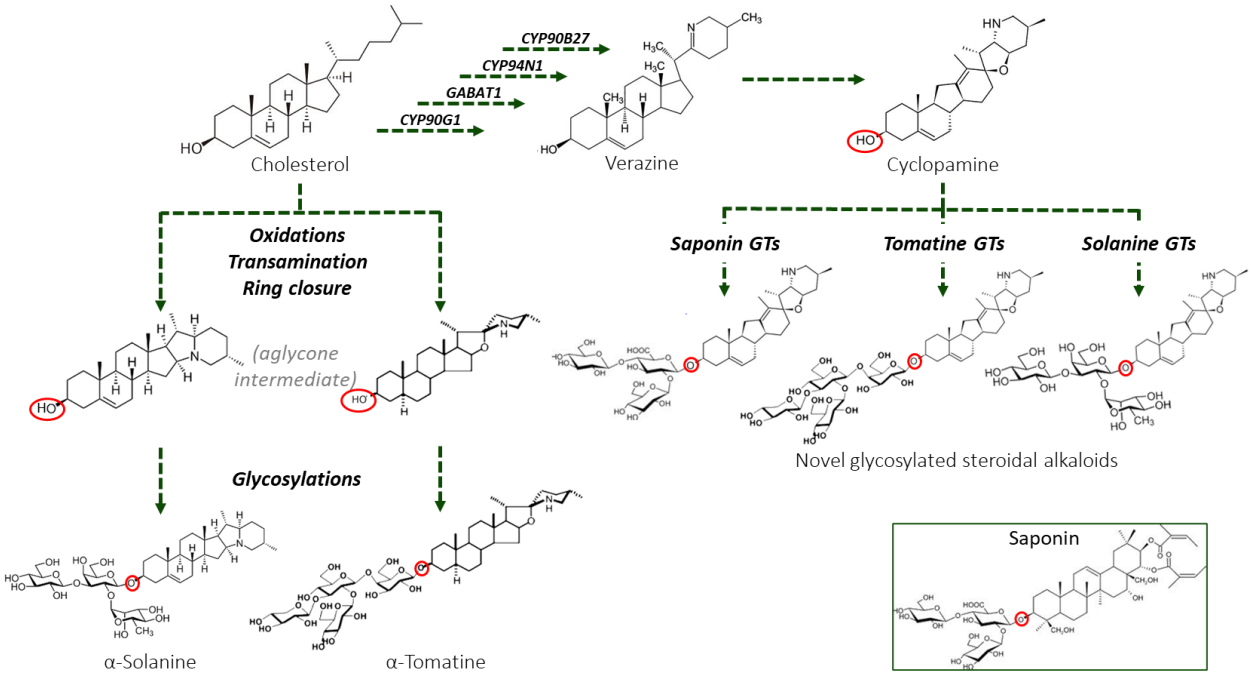


Figure 24. Potential of a cholesterol producing yeast platform for glycosylated steroidal alkaloid biosynthesis. Cholesterol is a primary substrate for numerous relevant plant natural products. As more details about specific pathways are discovered, there will be ample opportunity for implementing yeast as a chassis for the bioproduction of these molecules. There is additional capacity to produce novel glycosylated steroidal alkaloids with the integration of GTs from various pathways. Saponins, like many glycosylated steroidal alkaloids, are heavily glycosylated at the 3C position of the scaffold (highlighted in red). The promiscuous nature of GTs can potentially be co-opted by integrating both native and non-native GTs with varying scaffolds to produce new-to-nature molecules as well as economically or medically relevant natural glycosylated steroidal alkaloids.

.

Final thoughts

While much of the work discussed is varied in its approach or function, its core follows the reductionist theme of synthetic biology, aiming to rebuild living systems for specific applications in biotechnology. Whether engineering a fungus or a plant, the concepts applied remain the same. The amazing diversity of biological systems present on earth are a canvas for synthetic biologists, providing an almost infinite source of genetic information that can be reengineered and reshaped to solve challenging problems. In my opinion, synthetic biology is still in its infancy, but as we collectively increase our knowledge and understanding of the core tenets of life, we will see the true power of synthetic biology emerge. Even with these limits on our current understanding, we are transforming the old industrial economic model into a biobased economic model. Opportunities for ambitious and clever engineers are abundant, and I predict an exponential increase in the development of novel synthetic systems in the coming decade. So much untapped potential makes this an exciting era for biotechnology, with the prospect for application of bio-based solutions across all sectors. Eventually, the integration of biological systems with digital systems will exponentially expand the possibilities, as electrical and biological process can be intertwined to generate platforms we have yet to imagine, all made possible with synthetic biology.

References

1. Biermann, Frank, Norichika Kanie, and Rakhyun E. Kim. "Global governance by goal-setting: the novel approach of the UN Sustainable Development Goals." *Current Opinion in Environmental Sustainability* 26 (2017): 26-31.

2. Benner, Steven A., and A. Michael Sismour. "Synthetic biology." *Nature reviews genetics* 6.7 (2005): 533-543.

3. Ruder, Warren C., Ting Lu, and James J. Collins. "Synthetic biology moving into the clinic." *Science* 333.6047 (2011): 1248-1252.

4. Wurtzel, Eleanore T., et al. "Revolutionizing agriculture with synthetic biology." *Nature Plants* 5.12 (2019): 1207-1210.

5. Lv, Xueqin, et al. "Synthetic biology for future food: Research progress and future directions." *Future Foods* 3 (2021): 100025.

6. Rabinovitch-Deere, Christine A., et al. "Synthetic biology and metabolic engineering approaches to produce biofuels." *Chemical reviews* 113.7 (2013): 4611-4632.

7. Cravens, Aaron, James Payne, and Christina D. Smolke. "Synthetic biology strategies for microbial biosynthesis of plant natural products." *Nature communications* 10.1 (2019): 1-12.

8. Paddon, Chris J., and Jay D. Keasling. "Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development." *Nature reviews microbiology* 12.5 (2014): 355-367.

9. Tang, Tzu-Chieh, et al. "Materials design by synthetic biology." *Nature Reviews Materials* 6.4 (2021): 332-350.

10. Doudna, Jennifer A., and Emmanuelle Charpentier. "The new frontier of genome engineering with CRISPR-Cas9." *Science* 346.6213 (2014): 1258096.

11. Hughes, Randall A., and Andrew D. Ellington. "Synthetic DNA synthesis and assembly: putting the synthetic in synthetic biology." *Cold Spring Harbor perspectives in biology* 9.1 (2017): a023812.

12. Gibson, Daniel G., et al. "Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome." *science* 319.5867 (2008): 1215-1220.

13. Suyama, Yoshihisa, et al. *Complementary combination of multiplex high-throughput DNA sequencing for molecular phylogeny*. Vol. 37. No. 1. Hoboken, USA: John Wiley & Sons, Inc., 2022.

14. Schwab, Klaus. *The Fourth Industrial Revolution*. Portfolio Penguin, 2017.

15. Bouckaert, Stéphanie, et al. "Net Zero by 2050: A Roadmap for the Global Energy Sector." (2021).

16. US Environmental Protection Agency (2020). Inventory of U.S. Greenhouse Gas Emissions and Sinks. https://www.epa.gov/sites/production/files/2020-04/documents/us-ghg-inventory-2020-main-text.pdf.

17. US Department of Energy; Alternative fuels data center (2020a). Global ethanol production statistics. https://afdc.energy.gov/data/10331.

18. 109th US Congress (2005). Energy policy act of 2005.

19. Biotechnology Innovation Organization (BIO) (2014). The Renewable Fuel Standard: A Decade's worth of carbon reductions. https://www.bio.org/sites/default/files/legacy/bioorg/docs/RFS%2010%20Year%20GHG%

20. US Department of Energy; Energy Information Administration (U.S.E.I.A) (2020b). Monthly Energy Review. https://www.eia.gov/totalenergy/data/monthly/archive/00352003.pdf.

21. Isikgor, Furkan H., and C. Remzi Becer. "Lignocellulosic biomass: a sustainable platform for the production of bio-based chemicals and polymers." *Polymer Chemistry* 6.25 (2015): 4497-4559.
22. Field, John L., et al. "Robust paths to net greenhouse gas mitigation and negative emissions via advanced biofuels." *Proceedings of the National Academy of Sciences* 117.36 (2020): 21968-21977.
23. Tilman, David, Jason Hill, and Clarence Lehman. "Carbon-negative biofuels from low-input high-diversity grassland biomass." *Science* 314.5805 (2006): 1598-1600.
24. 110th US Congress (2007). Energy independence and security act of 2007.
25. Jiang, Dong, et al. "Potential bioethanol production from sweet sorghum on marginal land in China." *Journal of Cleaner Production* 220 (2019): 225-234.
26. Diallo, Boubacar, et al. "Biomass yield, chemical composition and theoretical ethanol yield for different genotypes of energy sorghum cultivated on marginal land in China." *Industrial Crops and Products* 137 (2019): 221-230.
27. Kromdijk, Johannes, et al. "Improving photosynthesis and crop productivity by accelerating recovery from photoprotection." *Science* 354.6314 (2016): 857-861.
28. Kirst, Henning, et al. "Photosynthetic antenna engineering to improve crop yields." *Planta* 245.5 (2017): 1009-1020.
29. Shih, Patrick M., et al. "Introduction of a synthetic CO2-fixing photorespiratory bypass into a cyanobacterium." *Journal of Biological Chemistry* 289.14 (2014): 9493-9500.
30. South, Paul F., et al. "Synthetic glycolate metabolism pathways stimulate crop growth and productivity in the field." *Science* 363.6422 (2019): eaat9077.
31. Belcher, Michael S., et al. "Design of orthogonal regulatory systems for modulating gene expression in plants." *Nature chemical biology* 16.8 (2020): 857-865.
32. Buyel, Johannes F. "Plant molecular farming–Integration and exploitation of side streams to achieve sustainable biomanufacturing." *Frontiers in plant science* 9 (2019): 1893.
33. Yang, Minliang, et al. "Accumulation of high-value bioproducts in planta can improve the economics of advanced biofuels." *Proceedings of the National Academy of Sciences* 117.15 (2020): 8639-8648.
34. Capell, Teresa, et al. "Potential applications of plant biotechnology against SARS-CoV-2." *Trends in plant science* 25.7 (2020): 635-643.
35. Dent, Matthew, and Nobuyuki Matoba. "Cancer biologics made in plants." *Current opinion in biotechnology* 61 (2020): 82-88.
36. Donini, Marcello, and Carla Marusic. "Current state-of-the-art in plant-based antibody production systems." *Biotechnology letters* 41.3 (2019): 335-346.
37. Mortimer, Jenny C. "Plant synthetic biology could drive a revolution in biofuels and medicine." *Experimental Biology and Medicine* 244.4 (2019): 323-331.
38. Clark, Michael, and Maciej Maselko. "Transgene biocontainment strategies for molecular farming." *Frontiers in Plant Science* 11 (2020): 210.
39. Baral, Nawa Raj, et al. "Approaches for more efficient biological conversion of lignocellulosic feedstocks to biofuels and bioproducts." *ACS sustainable chemistry & engineering* 7.10 (2019): 9062-9079.
40. Pimienta, José A. Pérez, et al. "Pilot-scale hydrothermal pretreatment and optimized saccharification enables bisabolene production from multiple feedstocks." *Green Chemistry* 21.11 (2019): 3152-3164.

41. Dos Santos, Antonio Carlos, et al. "Lignin–enzyme interactions in the hydrolysis of lignocellulosic biomass." *Trends in biotechnology* 37.5 (2019): 518-531.
42. Aznar, Aude, et al. "Gene stacking of multiple traits for high yield of fermentable sugars in plant biomass." *Biotechnology for biofuels* 11.1 (2018): 1-14.
43. Eudes, Aymerick, et al. "Expression of a bacterial 3-dehydroshikimate dehydratase reduces lignin content and improves biomass saccharification efficiency." *Plant biotechnology journal* 13.9 (2015): 1241-1250.
44. Yan, Jingwei, et al. "Increased drought tolerance in plants engineered for low lignin and low xylan content." *Biotechnology for biofuels* 11.1 (2018): 1-11.
45. Fang, Zhen, Matthew C. Weisenberger, and Mark S. Meier. "Utilization of lignin-derived small molecules: Epoxy polymers from lignin oxidation products." *ACS Applied Bio Materials* 3.2 (2020): 881-890.
46. Incha, Matthew R., et al. "Leveraging host metabolism for bisdemethoxycurcumin production in Pseudomonas putida." *Metabolic engineering communications* 10 (2020): e00119.
47. Wijffels, René H., and Maria J. Barbosa. "An outlook on microalgal biofuels." *Science* 329.5993 (2010): 796-799.
48. Alalwan, Hayder A., Alaa H. Alminshid, and Haydar AS Aljaafari. "Promising evolution of biofuel generations. Subject review." *Renewable Energy Focus* 28 (2019): 127-139.
49. Abdullah, Bawadi, et al. "Fourth generation biofuel: A review on risks and mitigation strategies." *Renewable and sustainable energy reviews* 107 (2019): 37-50.
50. Zhu, Yunhua, Susanne B. Jones, and Daniel B. Anderson. *Algae farm cost model: Considerations for photobioreactors*. No. PNNL-28201. Pacific Northwest National Lab.(PNNL), Richland, WA (United States), 2018.
51. Melis, Anastasios, John Neidhardt, and John R. Benemann. "Dunaliella salina (Chlorophyta) with small chlorophyll antenna sizes exhibit higher photosynthetic productivities and photon use efficiencies than normally pigmented cells." *Journal of Applied Phycology* 10.6 (1998): 515-525.
52. Melis, Anastasios. "Solar energy conversion efficiencies in photosynthesis: minimizing the chlorophyll antennae to maximize efficiency." *Plant science* 177.4 (2009): 272-280.
53. Blankenship, Robert E., et al. "Comparing photosynthetic and photovoltaic efficiencies and recognizing the potential for improvement." *science* 332.6031 (2011): 805-809.
54. Chen, Min, et al. "A red-shifted chlorophyll." *Science* 329.5997 (2010): 1318-1319.
55. Tros, Martijn, et al. "Harvesting far-red light: Functional integration of chlorophyll f into Photosystem I complexes of Synechococcus sp. PCC 7002." *Biochimica et Biophysica Acta (BBA)-Bioenergetics* 1861.8 (2020): 148206.
56. Savir, Yonatan, et al. "Cross-species analysis traces adaptation of Rubisco toward optimality in a low-dimensional landscape." *Proceedings of the National Academy of Sciences* 107.8 (2010): 3475-3480.
57. Liang, Feiyan, et al. "Engineered cyanobacteria with enhanced growth show increased ethanol production and higher biofuel to biomass ratio." *Metabolic Engineering* 46 (2018): 51-
58. Oliver, John WK, and Shota Atsumi. "A carbon sink pathway increases carbon productivity in cyanobacteria." *Metabolic engineering* 29 (2015): 106-112.
59. Claassens, Nico J., et al. "Harnessing the power of microbial autotrophy." *Nature Reviews Microbiology* 14.11 (2016): 692-706.
60. Schwander, Thomas, et al. "A synthetic pathway for the fixation of carbon dioxide in vitro." *Science* 354.6314 (2016): 900-904.

61. Alalwan, Hayder A., Alaa H. Alminshid, and Haydar AS Aljaafari. "Promising evolution of biofuel generations. Subject review." *Renewable Energy Focus* 28 (2019): 127-139.
62. De Bhowmick, Goldy, Lokanand Koduru, and Ramkrishna Sen. "Metabolic pathway engineering towards enhancing microalgal lipid biosynthesis for biofuel application—a review." *Renewable and Sustainable Energy Reviews* 50 (2015): 1239-1253.
63. Rashid, Naim, Muhammad Saif Ur Rehman, and Jong-In Han. "Recycling and reuse of spent microalgal biomass for sustainable biofuels." *Biochemical engineering journal* 75 (2013): 101-107.
64. Kings, Ajith J., et al. "Cultivation, extraction and optimization of biodiesel production from potential microalgae Euglena sanguinea using eco-friendly natural catalyst." *Energy Conversion and Management* 141 (2017): 224-235.
65. Knoot, Cory J., et al. "Cyanobacteria: promising biocatalysts for sustainable chemical production." *Journal of Biological Chemistry* 293.14 (2018): 5044-5052.
66. Gao, Zhengxu, et al. "Photosynthetic production of ethanol from carbon dioxide in genetically engineered cyanobacteria." *Energy & Environmental Science* 5.12 (2012): 9857-9865.
67. Liu, Xufeng, et al. "Modular engineering for efficient photosynthetic biosynthesis of 1-butanol from CO 2 in cyanobacteria." *Energy & Environmental Science* 12.9 (2019): 2765-2777.
68. Kopf, Matthias, et al. "Comparative analysis of the primary transcriptome of Synechocystis sp. PCC 6803." *DNA research* 21.5 (2014): 527-539.
69. Shabestary, Kiyan, and Elton P. Hudson. "Computational metabolic engineering strategies for growth-coupled biofuel production by Synechocystis." *Metabolic engineering communications* 3 (2016): 216-226.
70. Donald CM: The breeding of crop ideotypes. *Euphytica* (1968), 17:385–403.
71. Pauly M, Keegstra K: Cell-wall carbohydrates and their modification as a resource for biofuels. *Plant J* (2008), 54:559–568.
72. Himmel ME, Ding S-Y, Johnson DK, Adney WS, Nimlos MR, Brady JW, Foust TD: Biomass recalcitrance: engineering plants and enzymes for biofuels production. *Science* 2007, 315:804–807.
73. DeMartini JD, Pattathil S, Miller JS, Li H, Hahn MG, Wyman CE: Investigating plant cell wall components that affect biomass recalcitrance in poplar and switchgrass. *Energy & Environmental Science* 2013, 6:898.
74. Eudes A, Sathitsuksanoh N, Baidoo EEK, George A, Liang Y, Yang F, Singh S, Keasling JD, Simmons BA, Loqué D: Expression of a bacterial 3-dehydroshikimate dehydratase reduces lignin content and improves biomass saccharification efficiency. *Plant Biotechnol J* 2015, 13:1241–1250.
75. Petersen PD, Lau J, Ebert B, Yang F, Verhertbruggen Y, Kim JS, Varanasi P, Suttangkakul A, Auer M, Loqué D, et al.: Engineering of plants with improved properties as biofuels feedstocks by vessel-specific complementation of xylan biosynthesis mutants. *Biotechnol Biofuels* 2012, 5:84.
76. Scheller HV: Method of Reducing Acetylation in Plants to Improve Biofuel Production. *US Patent* 2012,
77. Aznar A, Chalvin C, Shih PM, Maimann M, Ebert B, Birdseye DS, Loqué D, Scheller HV: Gene stacking of multiple traits for high yield of fermentable sugars in plant biomass. *Biotechnol Biofuels* 2018, 11:2.

78. Kirst H, Gabilly ST, Niyogi KK, Lemaux PG, Melis A: Photosynthetic antenna engineering to improve crop yields. *Planta* 2017, 245:1009–1020.
79. Kromdijk J, Głowacka K, Leonelli L, Gabilly ST, Iwai M, Niyogi KK, Long SP: Improving photosynthesis and crop productivity by accelerating recovery from photoprotection. *Science* 2016, 354:857–861.
80. Shih PM, Zarzycki J, Niyogi KK, Kerfeld CA: Introduction of a Synthetic CO2-fixing Photorespiratory Bypass into a Cyanobacterium. *J Biol Chem* 2014, 289:9493–9500.
81. South PF, Cavanagh AP, Liu HW, Ort DR: Synthetic glycolate metabolism pathways stimulate crop growth and productivity in the field. *Science* 2019, 363.
82. Holwerda EK, Worthen RS, Kothari N, Lasky RC, Davison BH, Fu C, Wang Z-Y, Dixon RA, Biswal AK, Mohnen D, et al.: Multiple levers for overcoming the recalcitrance of lignocellulosic biomass. *Biotechnol Biofuels* 2019, 12:15.
83. Hassan SS, Williams GA, Jaiswal AK: Moving towards the second generation of lignocellulosic biorefineries in the EU: Drivers, challenges, and opportunities. *Renewable Sustainable Energy Rev* 2019, 101:590–599.
84. Huang K, Won W, Barnett KJ, Brentzel ZJ, Alonso DM, Huber GW, Dumesic JA, Maravelias CT: Improving economics of lignocellulosic biofuels: An integrated strategy for coproducing 1,5-pentanediol and ethanol. *Appl Energy* 2018, 213:585–594.
85. Vanhercke T, Belide S, Taylor MC, El Tahchy A, Okada S, Rolland V, Liu Q, Mitchell M, Shrestha P, Venables I, et al.: Up-regulation of lipid biosynthesis increases the oil content in leaves of Sorghum bicolor. *Plant Biotechnol J* 2019, 17:220–232.
86. Kalghatgi G: Is it really the end of internal combustion engines and petroleum in transport? *Appl Energy* 2018, 225:965–974.
87. Nair S, Paulose H: Emergence of green business models: The case of algae biofuel for aviation. *Energy Policy* 2014, 65:175–184.
88. Shonnard DR, Williams L, Kalnes TN: Camelina-derived jet fuel and diesel: Sustainable advanced biofuels. *Environ Prog Sustain Energy* 2010, 29:382–392.
89. Nandi S, Kwong AT, Holtz BR, Erwin RL, Marcel S, McDonald KA: Techno-economic analysis of a transient plant-based platform for monoclonal antibody production. *MAbs* 2016, 8:1456–1466.
90. Fuentes P, Zhou F, Erban A, Karcher D, Kopka J, Bock R: A new synthetic biology approach allows transfer of an entire metabolic pathway from a medicinal plant to a biomass crop. *Elife* 2016, 5.
91. Hunt VD: Gasohol handbook. 1981,
92. USDA ERS - U.S. Bioenergy Statistics. *USDA ERA* 2018,
93. Rosegrant MW: *Biofuels and grain prices: impacts and policy responses*. International Food Policy Research Institute Washington, DC; 2008.
94. German L, Goetz A, Searchinger T, Oliveira G de LT, Tomei J, Hunsberger C, Weigelt J: Sine Qua Nons of sustainable biofuels: Distilling implications of under-performance for national biofuel programs. *Energy Policy* 2017, 108:806–817.
95. Rathore D, Nizami A-S, Singh A, Pant D: Key issues in estimating energy and greenhouse gas savings of biofuels: challenges and perspectives. *Biofuel Research Journal* 2016, 3:380–393.
96. Fargione J, Hill J, Tilman D, Polasky S, Hawthorne P: Land clearing and the biofuel carbon debt. *Science* 2008, 319:1235–1238.

97. Rulli MC, Casirati S, Dell'Angelo J, Davis KF, Passera C, D'Odorico P: Interdependencies and telecoupling of oil palm expansion at the expense of Indonesian rainforest. *Renewable Sustainable Energy Rev* 2019, 105:499–512.
98. Budzianowski WM: High-value low-volume bioproducts coupled to bioenergies with potential to enhance business development of sustainable biorefineries. *Renewable Sustainable Energy Rev* 2017, 70:793–804.
99. Hybrid Poplar Composition - Idaho National Laboratory. *Bioenergy Feedstock Library - Idaho National Laboratory* 2018,
100. Switchgrass Composition - Idaho National Laboratory. *Bioenergy Library* 2011,
101. Sorghum Composition - Idaho National Laboratory. *Bioenergy Library* 2016,
102. Sugarcane Bagasse composition - Idaho National Laboratory. *Bioenergy Library* 2016,
103. Godin B, Agneessens R, Gerin PA, Delcarte J: Composition of structural carbohydrates in biomass: Precision of a liquid chromatography method using a neutral detergent extraction and a charged aerosol detector. *Talanta* 2011, 85:2014–2026.
104. Neutelings G: Lignin variability in plant cell walls: contribution of new models. *Plant Sci* 2011, 181:379–386.
105. He J, Ma C, Ma Y, Li H, Kang J, Liu T, Polle A, Peng C, Luo Z-B: Cadmium tolerance in six poplar species. *Environ Sci Pollut Res Int* 2013, 20:163–174.
106. Decker SR, Carlile M, Selig MJ, Doeppke C, Davis M, Sykes R, Turner G, Ziebell A: Reducing the effect of variable starch levels in biomass recalcitrance screening. *Methods Mol Biol* 2012, 908:181–195.
107. Zhao YL, Steinberger Y, Shi M, Han LP, Xie GH: Changes in stem composition and harvested produce of sweet sorghum during the period from maturity to a sequence of delayed harvest dates. *Biomass Bioenergy* 2012, 39:261–273.
108. Canilha, Larissa, et al. "Bioconversion of sugarcane biomass into ethanol: an overview about composition, pretreatment methods, detoxification of hydrolysates, enzymatic saccharification, and ethanol fermentation." *Journal of Biomedicine and Biotechnology* 2012 (2012).
109. Bagheri M, Mansouri H: Effect of induced polyploidy on some biochemical parameters in Cannabis sativa L. *Appl Biochem Biotechnol* 2015, 175:2366–2375.
110. Wyman CE, Balan V, Dale BE, Elander RT, Falls M, Hames B, Holtzapple MT, Ladisch MR, Lee YY, Mosier N, et al.: WITHDRAWN: Comparative Data on Effects of Leading Pretreatments and Enzyme Loadings and Formulations on Sugar Yields from Different Switchgrass Sources. *Bioresource Technology* 2011, doi:10.1016/j.biortech.2011.04.030.
111. Xu F, Zhou L, Zhang K, Yu J, Wang D: Rapid Determination of Both Structural Polysaccharides and Soluble Sugars in Sorghum Biomass Using Near-Infrared Spectroscopy. *Bioenergy Res* 2015, 8:130–136.
112. Lingle SE, Thomson JL: Sugarcane Internode Composition During Crop Development. *Bioenergy Res* 2012, 5:168–178.
113. Godin B, Lamaudière S, Agneessens R, Schmit T, Goffart J-P, Stilmant D, Gerin PA, Delcarte J: Chemical Composition and Biofuel Potentials of a Wide Diversity of Plant Biomasses. *Energy & Fuels* 2013, 27:2588–2598.
114. Brenneisen R: Chemistry and Analysis of Phytocannabinoids and Other Cannabis Constituents. In *Marijuana and the Cannabinoids*. Edited by ElSohly MA. Humana Press; 2007:17–49. (+114 to current cit)

115. Shih, P. M., Liang, Y. & Loqué, D. Biotechnology and synthetic biology approaches for metabolic engineering of bioenergy crops. *Plant J.* 87, 103–117 (2016).
116. Hawkins, Kristy M., and Christina D. Smolke. "Production of benzylisoquinoline alkaloids in Saccharomyces cerevisiae." *Nature chemical biology* 4.9 (2008): 564-573.
117. Potvin-Trottier, L., Lord, N. D., Vinnicombe, G. & Paulsson, J. Synchronous long-term oscillations in a synthetic gene circuit. *Nature* 538, 514–517 (2016).
118. Brückner, K. *et al.* A library of synthetic transcription activator-like effector-activated promoters for coordinated orthogonal gene expression in plants. *Plant J.* 82, 707–716 (2015).
119. Bashor, C. J. *et al.* Complex signal processing in synthetic gene circuits using cooperative regulatory assemblies. *Science* 364, 593–597 (2019).
120. Farzadfard, F., Perli, S. D. & Lu, T. K. Tunable and multifunctional eukaryotic transcription factors based on CRISPR/Cas. *ACS Synth. Biol.* 2, 604–613 (2013).
121. Lebar, T. & Jerala, R. Benchmarking of TALE- and CRISPR/dCas9-Based Transcriptional Regulators in Mammalian Cells for the Construction of Synthetic Genetic Circuits. *ACS Synth. Biol.* 5, 1050–1058 (2016).
122. Gaber, R. *et al.* Designable DNA-binding domains enable construction of logic circuits in mammalian cells. *Nat. Chem. Biol.* 10, 203–208 (2014).
123. Khalil, A. S. *et al.* A Synthetic Biology Framework for Programming Eukaryotic Transcription Functions. *Cell* 150 647–658 (2012).
124. Haseloff, J. GFP variants for multispectral imaging of living cells. *Methods Cell Biol.* 58, 139–151 (1999).
125. Laplaze, L. *et al.* GAL4-GFP enhancer trap lines for genetic manipulation of lateral root development in Arabidopsis thaliana. *J. Exp. Bot.* 56, 2433–2442 (2005).
126. Gardner, M. J. *et al.* GAL4 GFP enhancer trap lines for analysis of stomatal guard cell development and gene expression. *J. Exp. Bot.* 60, 213–226 (2009).
127. Smale, S. T. & Kadonaga, J. T. The RNA Polymerase II Core Promoter. *Annual Review of Biochemistry* 72 449–479 (2003).
128. Lubliner, S. *et al.* Core promoter sequence in yeast is a major determinant of expression level. *Genome Res.* 25, 1008–1017 (2015).
129. Pan, S., Czarnecka-Verner, E. & Gurley, W. B. Role of the TATA Binding Protein–Transcription Factor IIB Interaction in Supporting Basal and Activated Transcription in Plant Cells. *The Plant Cell* 12 125–135 (2000).
130. Horikoshi, M. *et al.* Transcription factor TFIID induces DNA bending upon binding to the TATA element. *Proc. Natl. Acad. Sci. U. S. A.* 89, 1060–1064 (1992).
131. Hampsey, M. Molecular Genetics of the RNA Polymerase II General Transcriptional Machinery. *Microbiology and Molecular Biology Reviews* 62 465–503 (1998).
132. Sadowski, I., Ma, J., Triezenberg, S. & Ptashne, M. GAL4-VP16 is an unusually potent transcriptional activator. *Nature* 335, 563–564 (1988).
133. Taylor, I. C., Workman, J. L., Schuetz, T. J. & Kingston, R. E. Facilitated binding of GAL4 and heat shock factor to nucleosomal templates: differential function of DNA-binding domains. *Genes Dev.* 5, 1285–1298 (1991).
134. Kiran, K. *et al.* The TATA-box sequence in the basal promoter contributes to determining light-dependent gene expression in plants. *Plant Physiol.* 142, 364–376 (2006).
135. Joshi, C. P. An inspection of the domain between putative TATA box and translation start site in 79 plant genes. *Nucleic Acids Res.* 15, 6643–6653 (1987).

136. Vaillant, I., Schubert, I., Tourmente, S. & Mathieu, O. MOM1 mediates DNA-methylation-independent silencing of repetitive sequences in Arabidopsis. *EMBO reports* 7 1273–1278 (2006).
137. Okamoto, H. & Hirochika, H. Silencing of transposable elements in plants. *Trends in Plant Science* 6 527–534 (2001).
138. Matzke, M. A., Mette, M. F. & Matzke, A. J. Transgene silencing by the host genome defense: implications for the evolution of epigenetic control mechanisms in plants and vertebrates. *Plant Mol. Biol.* 43, 401–415 (2000).
139. Kooter, J. M., Matzke, M. A. & Meyer, P. Listening to the silent genes: transgene silencing, gene regulation and pathogen control. *Trends Plant Sci.* 4, 340–347 (1999).
140. Mutalik, V. K. *et al.* Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat. Methods* 10, 354–360 (2013).
141. Jensen, P. R. & Hammer, K. The sequence of spacers between the consensus sequences modulates the strength of prokaryotic promoters. *Appl. Environ. Microbiol.* 64, 82–87 (1998).
142. Levo, M. & Segal, E. In pursuit of design principles of regulatory sequences. *Nat. Rev. Genet.* 15, 453–468 (2014).
143. Harcum, S. W. & Bentley, W. E. Heat-Shock and Stringent Responses Have Overlapping Protease Activity in Escherichia coli: Implications for Heterologous Protein Yield. *Applied Biochemistry and Biotechnology* 80 23–38 (1999).
144. Denancé, N., Sánchez-Vallet, A., Goffner, D. & Molina, A. Disease resistance or growth: the role of plant hormones in balancing immune responses and fitness costs. *Front. Plant Sci.* 4, 155 (2013).
145. Tian, D., Traw, M. B., Chen, J. Q., Kreitman, M. & Bergelson, J. Fitness costs of R-gene-mediated resistance in Arabidopsis thaliana. *Nature* 423, 74–77 (2003).
146. Fordyce, P. M. *et al.* De novo identification and biophysical characterization of transcription-factor binding sites with microfluidic affinity analysis. *Nat. Biotechnol.* 28, 970–975 (2010).
147. Baker, C. R., Tuch, B. B. & Johnson, A. D. Extensive DNA-binding specificity divergence of a conserved transcription regulator. *Proc. Natl. Acad. Sci. U. S. A.* 108, 7493–7498 (2011).
148. Baker, C. R., Booth, L. N., Sorrells, T. R. & Johnson, A. D. Protein modularity, cooperative binding, and hybrid regulatory states underlie transcriptional network diversification. *Cell* 151, 80–95 (2012).
149. Goff, S. A., Cone, K. C. & Fromm, M. E. Identification of functional domains in the maize transcriptional activator C1: comparison of wild-type and dominant inhibitor proteins. *Genes Dev.* 5, 298–309 (1991).
150. Dingwall, C. & Laskey, R. A. Nuclear targeting sequences—a consensus? *Trends Biochem. Sci.* 16, 478–481 (1991).
151. Li, T., Stark, M. R., Johnson, A. D. & Wolberger, C. Crystal Structure of the MATa1/MATα2 Homeodomain Heterodimer Bound to DNA. *Science* 270, 262–269 (1995).
152. Hiratsu, K., Matsui, K., Koyama, T. & Ohme-Takagi, M. Dominant repression of target genes by chimeric repressors that include the EAR motif, a repression domain, inArabidopsis. *The Plant Journal* 34 733–739 (2003).
153. Mead, J. *et al.* Interactions of the Mcm1 MADS Box Protein with Cofactors That Regulate Mating in Yeast. *Molecular and Cellular Biology* 22 4607–4621 (2002).
154. Tan, S. & Richmond, T. J. Crystal structure of the yeast MATα2/MCM1/DNA ternary complex. *Nature* 391 660–666 (1998).

155. Elble, R. & Tye, B. K. Both activation and repression of a-mating-type-specific genes in yeast require transcription factor Mcm1. *Proceedings of the National Academy of Sciences* 88 10966–10970 (1991).
156. Fagard, M. & Vaucheret, H. (TRANS)GENE SILENCING IN PLANTS: How Many Mechanisms? *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 51, 167–194 (2000).
157. Morel, J. B., Mourrain, P., Béclin, C. & Vaucheret, H. DNA methylation and chromatin structure affect transcriptional and post-transcriptional transgene silencing in Arabidopsis. *Curr. Biol.* 10, 1591–1594 (2000).
158. Matzke, M. A., Primig, M., Trnovsky, J. & Matzke, A. J. Reversible methylation and inactivation of marker genes in sequentially transformed tobacco plants. *EMBO J.* 8, 643–649 (1989).
159. Sunilkumar, G., Mohr, L., Lopata-Finch, E., Emani, C. & Rathore, K. S. Developmental and tissue-specific expression of CaMV 35S promoter in cotton as revealed by GFP. *Plant Mol. Biol.* 50, 463–474 (2002).
160. Liu, W. *et al.* Computational discovery of soybean promoter cis-regulatory elements for the construction of soybean cyst nematode-inducible synthetic promoters. *Plant Biotechnol. J.* 12, 1015–1026 (2014).
161. Haberle, Vanja, and Alexander Stark. "Eukaryotic core promoters and the functional basis of transcription initiation." *Nature reviews Molecular cell biology* 19.10 (2018): 621-637.
162. Wen, Ruhui, et al. "PAX6 upstream antisense RNA (PAUPAR) inhibits colorectal cancer progression through modulation of the microRNA (miR)-17-5p/zinc finger protein 750 (ZNF750) axis." *Bioengineered* 12.1 (2021): 3886-3899.
163. Jupe, Florian, et al. "The complex architecture and epigenomic impact of plant T-DNA insertions." *PLoS genetics* 15.1 (2019): e1007819.
164. Liu, Ziqing, et al. "Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector." *Scientific reports* 7.1 (2017): 1-9.
165. Jores, Tobias, et al. "Synthetic promoter designs enabled by a comprehensive analysis of plant core promoters." *Nature Plants* 7.6 (2021): 842-855.
166. Dansako, Tomoko, et al. "5′ Untranslated region of theHSP18. 2 gene contributes to efficient translation in plant cells." *Journal of bioscience and bioengineering* 95.1 (2003): 52-58.
167. Struhl, Kevin. "Helix-turn-helix, zinc-finger, and leucine-zipper motifs for eukaryotic transcriptional regulatory proteins." *Trends in biochemical sciences* 14.4 (1989): 137-140.
168. Haberle, Vanja, and Alexander Stark. "Eukaryotic core promoters and the functional basis of transcription initiation." *Nature reviews Molecular cell biology* 19.10 (2018): 621-637.
169. Schäffner, Anton R., and Jen Sheen. "Maize rbcS promoter activity depends on sequence elements not found in dicot rbcS promoters." *The Plant Cell* 3.9 (1991): 997-1012.
170. Hull, Roger, S. N. Covey, and P. Dale. "Genetically modified plants and the 35S promoter: assessing the risks and enhancing the debate." *Microbial Ecology in Health and Disease* 12.1 (2000): 1-5.
171. Crick, Francis. "Central dogma of molecular biology." *Nature* 227.5258 (1970): 561-563.
172. Prakasham, Reddy Shetty, et al. "Sorghum biomass: a novel renewable carbon source for industrial bioproducts." *Biofuels* 5.2 (2014): 159-174.
173. Tyree, Melvin T., and Martin H. Zimmermann. *Xylem structure and the ascent of sap*. Springer Science & Business Media, 2013.
174. Salih, A. A., et al. "Rooting, water uptake, and xylem structure adaptation to drought of two sorghum cultivars." *Crop Science* 39.1 (1999): 168-173.

175. Petersen, Pia Damm, et al. "Engineering of plants with improved properties as biofuels feedstocks by vessel-specific complementation of xylan biosynthesis mutants." *Biotechnology for Biofuels* 5.1 (2012): 1-19.
176. Mitsuda, Nobutaka, et al. "NAC transcription factors, NST1 and NST3, are key regulators of the formation of secondary walls in woody tissues of Arabidopsis." *The Plant Cell* 19.1 (2007): 270-280.
177. Chai, Maofeng, et al. "The NAC transcription factor OsSWN1 regulates secondary cell wall development in Oryza sativa." *Journal of Plant Biology* 58.1 (2015): 44-51.
178. Zeng, Wei, et al. "Asparagus IRX9, IRX10, and IRX14A are components of an active xylan backbone synthase complex that forms in the Golgi apparatus." *Plant Physiology* 171.1 (2016): 93-109.
179. Brandon, Andrew G., Devon S. Birdseye, and Henrik V. Scheller. "A dominant negative approach to reduce xylan in plants." *Plant biotechnology journal* 18.1 (2020): 5.
180. Sharma, Rita, et al. "Agrobacterium-mediated transient transformation of sorghum leaves for accelerating functional genomics and genome editing studies." *BMC research notes* 13.1 (2020): 1-7.
181. Pettolino, Filomena A., et al. "Determining the polysaccharide composition of plant cell walls." *Nature protocols* 7.9 (2012): 1590-1607.
182. Barnes, William J., and Charles T. Anderson. "Acetyl bromide soluble lignin (ABSL) assay for total lignin quantification from plant biomass." *Bio-protocol* 7.5 (2017): e2149-e2149.
183. Petrovska, Biljana Bauer. "Historical review of medicinal plants' usage." *Pharmacognosy reviews* 6.11 (2012): 1.
184. McChesney, James D., Sylesh K. Venkataraman, and John T. Henri. "Plant natural products: back to the future or into extinction?." *Phytochemistry* 68.14 (2007): 2015-2022.
185. Debnath, Bikash, et al. "Role of plant alkaloids on human health: A review of biological activities." *Materials today chemistry* 9 (2018): 56-72.
186. Langenheim, Jean H. "Higher plant terpenoids: a phytocentric overview of their ecological roles." *Journal of chemical ecology* 20.6 (1994): 1223-1280.
187. Flores-Sanchez, Isvett J., and Robert Verpoorte. "Plant polyketide synthases: a fascinating group of enzymes." *Plant Physiology and Biochemistry* 47.3 (2009): 167-174.
188. Chen, Ruibing, et al. "Advanced strategies for production of natural products in yeast." *Iscience* 23.3 (2020): 100879.
189. De Montellano, Paul R. Ortiz, ed. *Cytochrome P450: structure, mechanism, and biochemistry*. Vol. 115. New York: Kluwer Academic/Plenum Publishers, 2005.
190. Connolly JD, Hill RA: Dictionary of Terpenoids. 1991, doi:10.1007/978-1-4899-4513-6.
192. Yazaki, Kazufumi, Gen-ichiro Arimura, and Toshiyuki Ohnishi. "'Hidden'terpenoids in plants: their biosynthesis, localization and ecological roles." *Plant and Cell Physiology* 58.10 (2017): 1615-1621.
193. Urlacher, Vlada B., and Marco Girhard. "Cytochrome P450 monooxygenases in biotechnology and synthetic biology." *Trends in biotechnology* 37.8 (2019): 882-897.
194. Valliere, Meaghan A., et al. "A cell-free platform for the prenylation of natural products and application to cannabinoid production." *Nature communications* 10.1 (2019): 1-9.
195. Booth, Judith K., Jonathan E. Page, and Jörg Bohlmann. "Terpene synthases from Cannabis sativa." *Plos one* 12.3 (2017): e0173911.
196. Allen, Keith D., et al. "Genomic characterization of the complete terpene synthase gene family from Cannabis sativa." *PloS one* 14.9 (2019): e0222363.

197. Braich, Shivraj, et al. "Generation of a comprehensive transcriptome atlas and transcriptome dynamics in medicinal cannabis." *Scientific reports* 9.1 (2019): 1-12.
198. Zager, Jordan J., et al. "Gene networks underlying cannabinoid and terpenoid accumulation in cannabis." *Plant physiology* 180.4 (2019): 1877-1897.
199. Rea, Kevin A., et al. "Biosynthesis of cannflavins A and B from Cannabis sativa L." *Phytochemistry* 164 (2019): 162-171.
200. Ban, Zhaonan, et al. "Noncatalytic chalcone isomerase-fold proteins in Humulus lupulus are auxiliary components in prenylated flavonoid biosynthesis." *Proceedings of the National Academy of Sciences* 115.22 (2018): E5223-E5232.
201. Esposito, Daniel, et al. "MaveDB: an open-source platform to distribute and interpret data from multiplexed assays of variant effect." *Genome biology* 20.1 (2019): 1-11.
202. Denby, Charles M., et al. "Industrial brewing yeast engineered for the production of primary flavor determinants in hopped beer." *Nature communications* 9.1 (2018): 1-10.
203. Lee, Michael E., et al. "A highly characterized yeast toolkit for modular, multipart assembly." *ACS synthetic biology* 4.9 (2015): 975-986.
204. Reider Apel, Amanda, et al. "A Cas9-based toolkit to program gene expression in Saccharomyces cerevisiae." *Nucleic acids research* 45.1 (2017): 496-508.
205. Luo, Xiaozhou, et al. "Complete biosynthesis of cannabinoids and their unnatural analogues in yeast." *Nature* 567.7746 (2019): 123-126.
206. Welling, Matthew T., et al. "Complex patterns of cannabinoid alkyl side-chain inheritance in cannabis." *Scientific reports* 9.1 (2019): 1-13.
207. Xu, Jinkun, et al. "Converting S-limonene synthase to pinene or phellandrene synthases reveals the plasticity of the active site." *Phytochemistry* 137 (2017): 34-41.
208. Saugspier, Michael, et al. "Hop bitter acids exhibit anti-fibrogenic effects on hepatic stellate cells in vitro." *Experimental and molecular pathology* 92.2 (2012): 222-228.
209. Levisson, Mark, et al. "Toward developing a yeast cell factory for the production of prenylated flavonoids." *Journal of agricultural and food chemistry* 67.49 (2019): 13478-13486.
210. Zirpel, Bastian, et al. "Optimization of Δ9-tetrahydrocannabinolic acid synthase production in Komagataella phaffii via post-translational bottleneck identification." *Journal of biotechnology* 272 (2018): 40-47.
211. Bathe U, Tissier A: Cytochrome P450 enzymes: A driving force of plant diterpene diversity. Phytochemistry 2019, 161:149–162.
212. Mitchell, Andrew J., and Jing-Ke Weng. "Unleashing the synthetic power of plant oxygenases: from mechanism to application." *Plant Physiology* 179.3 (2019): 813-829.
213. Song, Wei, et al. "Functional characterization of squalene epoxidase and NADPH-cytochrome P450 reductase in Dioscorea zingiberensis." *Biochemical and biophysical research communications* 509.3 (2019): 822-827.
214. Christensen, Ulla, et al. "De-bugging and maximizing plant cytochrome P450 production in Escherichia coli with C-terminal GFP fusions." *Applied microbiology and biotechnology* 101.10 (2017): 4103-4113.
215. Yang, Han, et al. "Reconstructing biosynthetic pathway of the plant-derived cancer chemopreventive-precursor glucoraphanin in Escherichia coli." *ACS Synthetic Biology* 7.1 (2018): 121-131.
216. Liu, Feixia, et al. "Biosynthesis of the high-value plant secondary product benzyl isothiocyanate via functional expression of multiple heterologous enzymes in Escherichia coli." *ACS synthetic biology* 5.12 (2016): 1557-1565.

217. Sevrioukova, Irina F., et al. "Structure of a cytochrome P450–redox partner electron-transfer complex." *Proceedings of the National Academy of Sciences* 96.5 (1999): 1863-1868.
218. Wong, Jeff, et al. "High-titer production of lathyrane diterpenoids from sugar by engineered Saccharomyces cerevisiae." *Metabolic engineering* 45 (2018): 142-148.
219. Ro, Dae-Kyun, et al. "Production of the antimalarial drug precursor artemisinic acid in engineered yeast." *Nature* 440.7086 (2006): 940-943.
220. Song, Chuankui, et al. "Attractive but toxic: emerging roles of glycosidically bound volatiles and glycosyltransferases involved in their formation." *Molecular Plant* 11.10 (2018): 1225-1236.
221. Augustin, Jörg M., et al. "Molecular activities, biosynthesis and evolution of triterpenoid saponins." *Phytochemistry* 72.6 (2011)
222. Takemura, Miho, Rie Tanaka, and Norihiko Misawa. "Pathway engineering for the production of β-amyrin and cycloartenol in Escherichia coli—a method to biosynthesize plant-derived triterpene skeletons in E. coli." *Applied microbiology and biotechnology* 101.17 (2017): 6615-6625.
223. Zhao, Yu-jia, and Chun Li. "Biosynthesis of plant triterpenoid saponins in microbial cell factories." *Journal of agricultural and food chemistry* 66.46 (2018): 12155-12165.
224. Moses, Tessa, et al. "Comparative analysis of CYP93E proteins for improved microbial synthesis of plant triterpenoids." *Phytochemistry* 108 (2014): 47-56.
225. Sun, Wentao, et al. "Novel trends for producing plant triterpenoids in yeast." *Critical reviews in biotechnology* 39.5 (2019): 618-632.
226. Wang, Pingping, et al. "Synthesizing ginsenoside Rh2 in Saccharomyces cerevisiae cell factory at high-efficiency." *Cell discovery* 5.1 (2019): 1-14.
227. Hardman, J. M., R. T. Brooke, and B. J. Zipp. "Cannabinoid glycosides: In vitro production of a new class of cannabinoids with improved physicochemical properties." *BioRxiv* (2017).
228. Bar-Peled, Maor, and Malcolm A. O'Neill. "Plant nucleotide sugar formation, interconversion, and salvage by sugar recycling." *Annual review of plant biology* 62.1 (2011): 127-155.
229. Kim, Jae-Eung, et al. "Tailoring the Saccharomyces cerevisiae endoplasmic reticulum for functional assembly of terpene synthesis pathway." *Metabolic engineering* 56 (2019): 50-59.
230. Arendt, Philipp, et al. "An endoplasmic reticulum-engineered yeast platform for overproduction of triterpenoids." *Metabolic engineering* 40 (2017): 165-175.
231. Zargar, Amin, et al. "Leveraging microbial biosynthetic pathways for the generation of 'drop-in'biofuels." *Current opinion in biotechnology* 45 (2017): 156-163.
232. Zhao, Cheng, et al. "Co-compartmentation of terpene biosynthesis and storage via synthetic droplet." *ACS synthetic biology* 7.3 (2018): 774-781.
233. Sadre, Radin, et al. "Cytosolic lipid droplets as engineered organelles for production and accumulation of terpenoid biomaterials in leaves." *Nature communications* 10.1 (2019): 1-12.
234. Ma, Tian, et al. "Lipid engineering combined with systematic metabolic engineering of Saccharomyces cerevisiae for high-yield production of lycopene." *Metabolic engineering* 52 (2019): 134-142.
235. Ma, Yi-Rong, et al. "Advances in the metabolic engineering of Yarrowia lipolytica for the production of terpenoids." *Bioresource technology* 281 (2019): 449-456.
236. Liu Y, Jiang X, Cui Z, Wang Z, Qi Q, Hou J: Engineering the oleaginous yeast Yarrowia lipolytica for production of α-farnesene. Biotechnology for Biofuels 2019, 12.

237. Liu, Guo-Song, et al. "The yeast peroxisome: a dynamic storage depot and subcellular factory for squalene overproduction." *Metabolic engineering* 57 (2020): 151-161.
238. Carvalho, Ângela, et al. "Designing microorganisms for heterologous biosynthesis of cannabinoids." *FEMS yeast research* 17.4 (2017).
239. Pandey RP: Diversifying Natural Products with Promiscuous Glycosyltransferase Enzymes via a Sustainable Microbial Fermentation Approach. Frontiers in Chemistry 2017, 5.
240. Martin, Virginia. "Overview of paclitaxel (Taxol®)." *Seminars in Oncology nursing*. Vol. 9. No. 4. WB Saunders, 1993.
241. Bhuju, Sudina, and Dhurva P. Gauchan. "Taxus wallichiana (Zucc.), an endangered anti-cancerous plant: a review." *Int J Res* 5.21 (2018): 10-21.
242. Zhang, C. H., et al. "Enhanced paclitaxel production induced by the combination of elicitors in cell suspension cultures of Taxus chinensis." *Biotechnology Letters* 22.19 (2000): 1561-1564.
243. Incardona, John P., et al. "The teratogenic Veratrum alkaloid cyclopamine inhibits sonic hedgehog signal transduction." *Development* 125.18 (1998): 3553-3562.
244. Horwitz, Susan Band, et al. "Taxol: mechanisms of action and resistance." *J. Natl. Cancer Inst. Monogr* (1992): 55-61.
245. Böttger, Angelika, et al. "Terpenes and terpenoids." *Lessons on caffeine, cannabis & Co*. Springer, Cham, 2018. 153-170.
246. Walls, Laura E., et al. "Optimizing the biosynthesis of oxygenated and acetylated Taxol precursors in Saccharomyces cerevisiae using advanced bioprocessing strategies." *Biotechnology and Bioengineering* 118.1 (2021): 279-293.
247. Lewis, David FV, and Peter Hlavica. "Interactions between redox partners in various cytochrome P450 systems: functional and structural aspects." *Biochimica et Biophysica Acta (BBA)-Bioenergetics* 1460.2-3 (2000): 353-374.
248. Zhang, Youjun, and Alisdair R. Fernie. "Metabolons, enzyme–enzyme assemblies that mediate substrate channeling, and their roles in plant metabolism." *Plant communications* 2.1 (2021): 100081.
249. McElroy, Christopher, and Stefan Jennewein. "Taxol® biosynthesis and production: from forests to fermenters." *Biotechnology of natural products*. Springer, Cham, 2018. 145-185.
250. Nowrouzi, Behnaz, et al. "Enhanced production of taxadiene in Saccharomyces cerevisiae." *Microbial cell factories* 19.1 (2020): 1-12.
251. Walker, Kevin, and Rodney Croteau. "Taxol biosynthetic genes." *Phytochemistry* 58.1 (2001): 1-7.
252. Gengenbach, Benjamin B., et al. "Comparison of microbial and transient expression (tobacco plants and plant-cell packs) for the production and purification of the anticancer mistletoe lectin viscumin." *Biotechnology and bioengineering* 116.9 (2019): 2236-2249.
253. Gengenbach, Benjamin B., et al. "Comparison of microbial and transient expression (tobacco plants and plant-cell packs) for the production and purification of the anticancer mistletoe lectin viscumin." *Biotechnology and bioengineering* 116.9 (2019): 2236-2249.
254. Pool, Martin R. "Signal recognition particles in chloroplasts, bacteria, yeast and mammals." *Molecular membrane biology* 22.1-2 (2005): 3-15.
255. Obata, Toshihiro. "Metabolons in plant primary and secondary metabolism." *Phytochemistry Reviews* 18.6 (2019): 1483-1507.
256. Bassard, Jean-Etienne, Birger Lindberg Møller, and Tomas Laursen. "Assembly of dynamic P450-mediated metabolons—order versus chaos." *Current Molecular Biology Reports* 3.1 (2017): 37-51.

257. Das, Aditi, and Stephen G. Sligar. "Modulation of the cytochrome P450 reductase redox potential by the phospholipid bilayer." *Biochemistry* 48.51 (2009): 12104-12112.
258. Pompon, Denis, et al. "[6] Yeast expression of animal and plant P450s in optimized redox environments." *Methods in enzymology*. Vol. 272. Academic Press, 1996. 51-64.
259. Van Roessel, Peter, and Andrea H. Brand. "Imaging into the future: visualizing gene expression and protein interactions with fluorescent proteins." *Nature Cell Biology* 4.1 (2002).
260. Pronk, Jack T. "Auxotrophic yeast strains in fundamental and applied research." *Applied and environmental microbiology* 68.5 (2002): 2095-2100.
261. Petriv, I., et al. "A new definition for the consensus sequence of the peroxisome targeting signal type 2." *Journal of molecular biology* 341.1 (2004): 119-134.
262. Augustin, Jörg M., et al. "Molecular activities, biosynthesis and evolution of triterpenoid saponins." *Phytochemistry* 72.6 (2011): 435-457.
263. Kärgel, Eva, et al. "Candida maltosa NADPH-cytochrome P450 reductase: cloning of a full-length cDNA, heterologous expression in Saccharomyces cerevisiae and function of the N-terminal region for membrane anchoring and proliferation of the endoplasmic reticulum." *Yeast* 12.4 (1996): 333-348.
264. Chen, Xiaoying, Jennica L. Zaro, and Wei-Chiang Shen. "Fusion protein linkers: property, design and functionality." *Advanced drug delivery reviews* 65.10 (2013): 1357-1369.
265. Gou, Mingyue, et al. "The scaffold proteins of lignin biosynthetic cytochrome P450 enzymes." *Nature plants* 4.5 (2018): 299-310.
266. Rahman, Atta-ur, and M. Iqbal Choudhary. "Chemistry and biology of steroidal alkaloids." *The Alkaloids: Chemistry and Biology*. Vol. 50. Academic Press, 1998. 61-108.
267. Souza, Cleiton M., et al. "A stable yeast strain efficiently producing cholesterol instead of ergosterol is functional for tryptophan uptake, but not weak organic acid resistance." *Metabolic engineering* 13.5 (2011): 555-569.
268. Clapp, Caitlin, et al. "Untangling the roles of anti-apoptosis in regulating programmed cell death using humanized yeast cells." *Frontiers in oncology* 2 (2012): 59.
269. Nijkamp, Jurgen F., et al. "De novo sequencing, assembly and analysis of the genome of the laboratory strain Saccharomyces cerevisiae CEN. PK113-7D, a model for modern industrial biotechnology." *Microbial cell factories* 11.1 (2012): 1-17.
270. Arneson, P. A., and R. D. Durbin. "Studies on the mode of action of tomatine as a fungitoxic agent." *Plant Physiology* 43.5 (1968): 683-686.
271. Fewell, Alison M., and James G. Roddick. "Interactive antifungal activity of the glycoalkaloids α-solanine and α-chaconine." *Phytochemistry* 33.2 (1993): 323-328.
272. Blankemeyer, James T., et al. "Effect of potato glycoalkaloids,. alpha.-chaconine and. alpha.-solanine on membrane potential of frog embryos." *Journal of Agricultural and Food Chemistry* 40.10 (1992): 2022-2025.
273. Augustin, Megan M., et al. "Elucidating steroid alkaloid biosynthesis in Veratrum californicum: production of verazine in Sf9 cells." *The Plant Journal* 82.6 (2015): 991-1003.
274. Gibeaut, David M. "Nucleotide sugars and glycosyltransferases for synthesis of cell wall matrix polysaccharides." *Plant Physiology and Biochemistry* 38.1-2 (2000): 69-80.
275. Xia, Hui, et al. "Recent developments in promiscuous enzymatic reactions for carbon-nitrogen bond formation." *Bioorganic Chemistry* (2022): 106014.
276. Kohara, Atsuko, et al. "Characterization and engineering of glycosyltransferases responsible for steroid saponin biosynthesis in Solanaceous plants." *Phytochemistry* 68.4 (2007): 478-486.