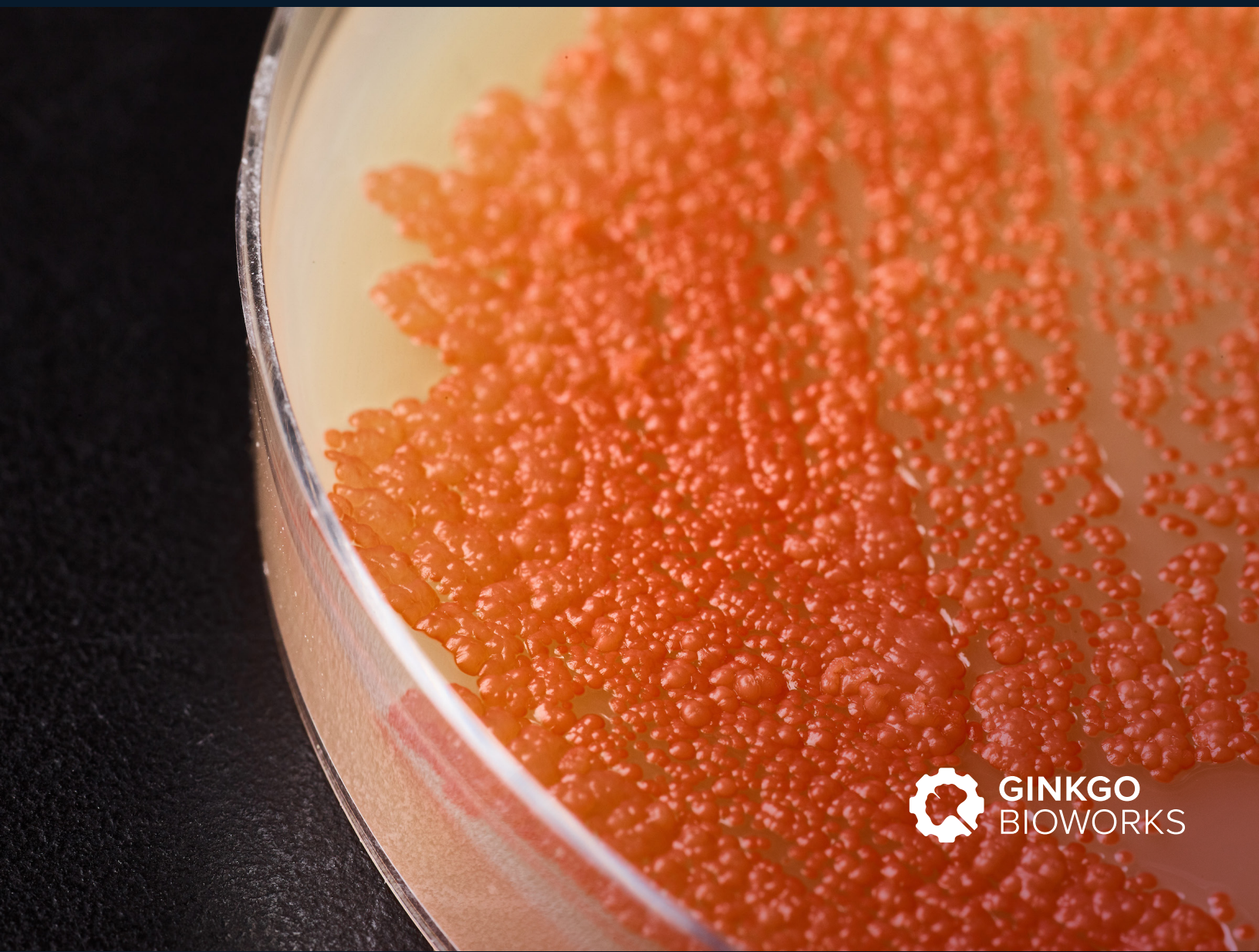
Bioproduction of Aromatic Molecules: Engineering Biology Towards Economic Viability

Deeya Burman, Donovan Layton, PhD, Nadia Parachin, PhD, Bernardo da Costa, PhD, Scott Marr, Gabriel Rodriguez, PhD, Ryan Morhard, Bruce Hilman

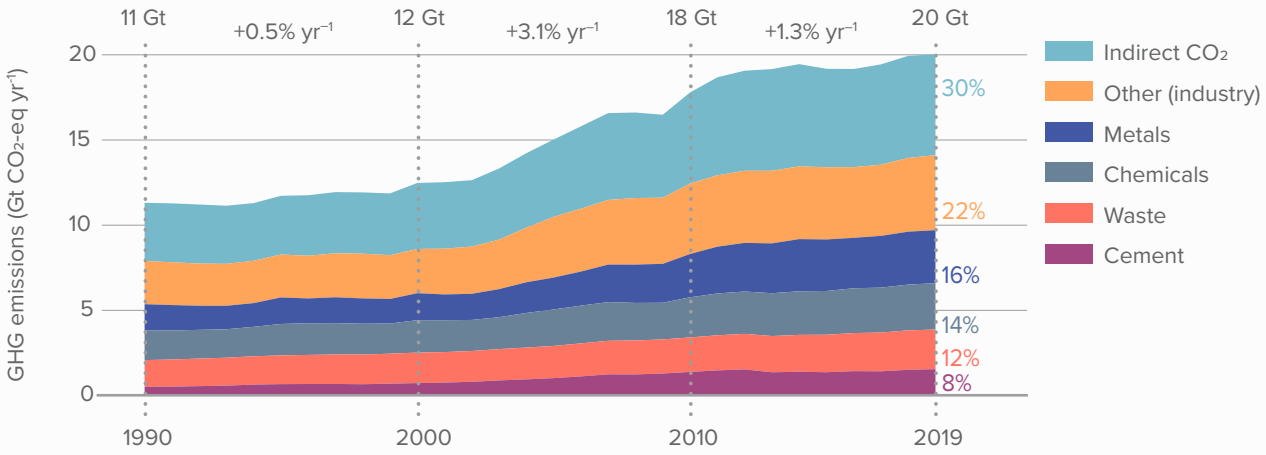


Background

Aromatic compounds are widely used and versatile in their applications. Historically, the primary source of aromatic compounds was coal tar. Post-World War II, petroleum became the dominant source of these compounds. Through a process referred to in the petroleum industry as "catalytic re-forming" or "hydroforming," petroleum, composed primarily of alkanes, is transformed into aromatic compounds. In recent years, questions have arisen around the impact of carbon intensity from petrochemical processes.

The chemicals sector accounts for 14% of industrial GHG emissions, with petrochemical production of high value chemicals accounting for ~2% of global emissions. Furthermore, experts anticipate that in the coming years, petrochemicals are expected to be the largest driver of world oil demand. In response, chemical companies have made concerted efforts to explore alternative, renewable feedstocks through R&D investment, strategic partnerships, and acquisitions.

Industry global GHG emission trends by subsectors

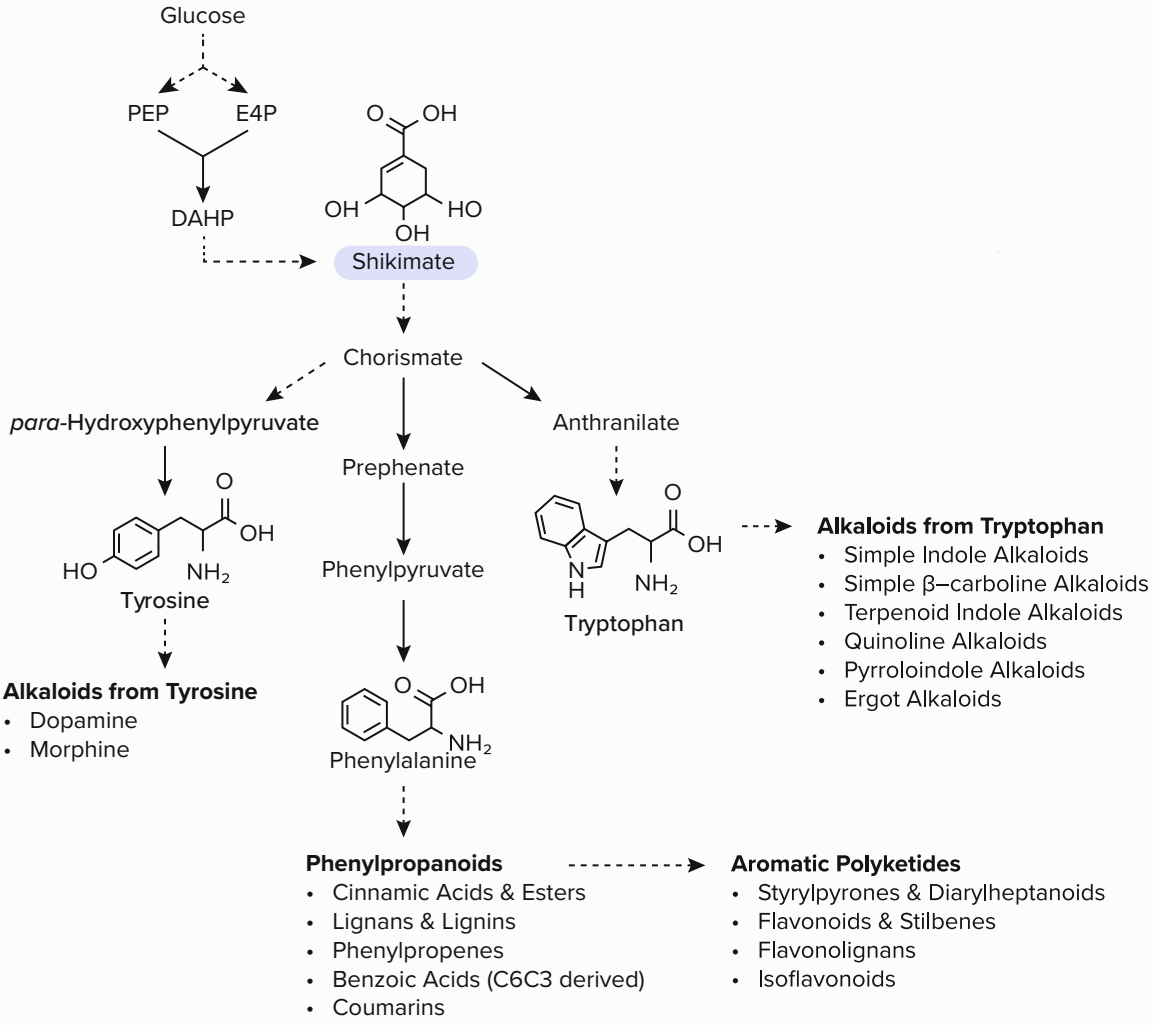


Source: IPCC Working Group III Summary for Policymakers

Bio-based production of aromatic chemicals has garnered significant interest by academic researchers and industry researchers alike. Companies including Ajinomoto, CJ CheilJedang, Evonik and Lesaffre have brought bio-based aromatic molecules to market in recent years.

Many of the natural metabolic pathways to synthesize aromatic compounds branch from a core pathway that produces shikimate. The shikimate pathway – the primary production route for aromatic amino acids, L-phenylalanine, L-tyrosine, and L-tryptophan – has been deeply studied in model organisms such as *Eschericia coli*, *Saccharomyces cerevisiae*, and *Arabidopsis thaliana*. Decades of academic biological research have elucidated the enzymatic steps along branches that advance from shikimate to yield flavonoids and other commercially relevant molecules. Recent advances that leverage synthetic biology approaches have further enabled engineered organisms to economically produce molecules using precision fermentation processes. The rate of progress continues to accelerate as new tools like artificial intelligence are implemented.

The Shikimate Pathway and Select Branches



The shikimate pathway comprises seven steps that start with intermediates from central metabolism in microbes, plants, and algae, specifically erythrose-4-phosphate (E4P) from the pentose phosphate pathway and phosphoenolpyruvate (PEP) from glycolysis. From there, a set of enzymes (or in *S. cerevisiae*, a penta-functional protein complex), catalyzes the transformation to shikimate, the aromatic precursor to protocatechuate and aminobenzoate and the first branch point in this pathway. From this branch point, a few additional enzymatic steps yield chorismate, the precursor to aromatic amino acids. Each shikimate branch point acts as a starting point for a diverse range of commerically relevant molecules.

To illustrate this diversity, consider the chorismate branches pictured above. In addition to aromatic amino acids, enzymatic pathways could yield organic acids like muconate and mandelate and stillbenoids like resveratrol in certain host organisms. The production rate of any of these molecules depends on a variety of factors – feedstock abundance, the specificity and activity of catalyzing enzymes and their abundance, toxicity of the molecule, and any intermediates produced along the way. Furthermore, enzymes along these pathways may be regulated by other molecules produced by the cell or may utilize a substrate that feeds competing pathways.

Synthetic Biology Approaches Have Progressed the Field

The complexity of gene expression and regulation along the shikimate pathway opens the door for synthetic biology approaches. Advances in synthetic biology, from DNA read/write, to protein, pathway, and cell systems engineering, through fermentation optimization and scale-up enable iterative design, build, and test approaches in complex systems. For example, DNA sequences corresponding to different promoters, repressors, or enzymes from other species can be tested in various combinations until the pathway in the desired host is optimized. Modern genetic engineering tools enable biological engineers to quickly test gene knockdowns, knockouts, and overexpression across competing pathway branches.

As mentioned earlier, the core shikimate pathway components have been well-studied, especially in model organisms such as *S. cerevisiae*. Branch points that lead to other industrially relevant precursors have also been elucidated in recent years. In parallel, synthetic biology researchers have characterized regulatory elements and feedback mechanisms to fine-tune genetic circuits in fermentation-friendly host organisms. A recent review paper demonstrates the culmination of this work: academic research employing synthetic biology and metabolic engineering methods have made significant strides toward commercially viable production of aromatic compounds.

Best-in-Class Titers Reported in the Literature as of 2018
Summary table from Averesch and Kromer, 2018

Target compound	Organism/strain	Carbon-source	Final titer (g/L)	Productivity (mg/L × h)	Relative carbon-yield (%)	Reference
Shikimate	*Corynebacterium glutamicum*	Glucose	141	2,937.5	66.18	Kogure et al., 2016
	Pichia stipitis	Glucose	3.11	25.92	14.96	Gao et al., 2017
cis,*cis*-Muconate	*Escherichia coli*	Glucose	36.8	766.67	26.13	Draths and Frost, 1994; Niu et al., 2002
	E. coli	Glucose + xylose (2:1)	4.7	65.28	52.95	Zhang et al., 2015
	E. coli	Glycerol + glucose	0.39	12.19	5.81	Sun et al., 2013
	E. coli	Glycerol + glucose	0.48	10	6.07	Sun et al., 2014
	E. coli	Glucose	0.605	8.4	10.99	Wang and Zheng, 2015
	E. coli	Glycerol + glucose	1.45	30.28	20.82	Lin et al., 2014b
	E. coli	Glucose	0.17	2.36	3.09	Sengupta et al., 2015
	Pseudomonas putida	p-Coumarate	15.6	213.7	101	Johnson et al., 2016
	P. putida	Glucose	4.92	91.11	10.29	Johnson et al., 2016
	Saccharomyces cerevisiae	Glucose	0.00156	0.0091765	0.0117	Weber et al., 2012
	S. cerevisiae	Glucose	0.14	1.31	10.6	Curran et al., 2013
	S. cerevisiae	Glucose	2.1	8.75	1.93	Leavitt et al., 2017
	S. cerevisiae	Glucose	0.32	4.45	1.2	Suástegui et al., 2017
Catechol	*E. coli*	Glucose	0.63	7.33	8.67	Pugh et al., 2014

Target compound	Organism/strain	Carbon-source	Final titer (g/L)	Productivity (mg/L × h)	Relative carbon-yield (%)	Reference
para-Hydroxybenzoate	*Klebsiella pneumonia*	Glucose	0.14	10.38	29.34	Müller et al., 1995
	E. coli	Glucose	12	23.96	19.92	Zhang et al., 2015
	Corynebacterium glutamicum	Glucose	36.6	1,527.38	65.85	Kitade et al., 2018
	P. putida	Glucose	1.73	54.06	23.66	Yu et al., 2016
	P. putida	Glucose	0.32	10.59	19.2	Verhoef et al., 2010
	P. putida	Glycerol	0.23	n.a.	23.51	Meijnen et al., 2011
	P. putida	Glucose + xylose (1:1)	0.18	0.19	19.85	Meijnen et al., 2011
	Saccharomyces cerevisiae	Glucose	0.09	0.82	1.12	Krömer et al., 2013
	S. cerevisiae	Glucose	0.15	2.08	0.92	Williams et al., 2015
	S. cerevisiae	Glucose	2.9	29	4.06	Averesch et al., 2017
Gallate	*E. coli*	Glucose + glycerol + yeast extract	1.27	35.17	n.d.	Chen et al., 2017
para-Aminobenzoate	*E. coli*	Glucose	4.8	100	32.31	Koma et al., 2014
	Corynebacterium glutamicum	Glucose	43	897.12	28.73	Kubota et al., 2016
	S. cerevisiae	Glucose	0.03	0.19	0.45	Krömer et al., 2013
	S. cerevisiae	Glycerol + ethanol	0.22	2.09	3.02	Averesch et al., 2016
meta-Aminobenzoate	*E. coli*	Glucose	0.048	0.333	0.404	Zhang and Stephanopoulos, 2016
ortho-Aminobenzoate	*P. putida*	Glucose	1.54	22.65	7.66	Kuepper et al., 2015
Tryptophan	*E. coli*	Glucose	40.2	1,005	32.82	Shen et al., 2012
R-Mandelate	*E. coli*	Glucose	0.68	28.33	16.32	Sun et al., 2011
Phenylethanol	*Kluyveromyces marxianus*	Glucose	1.3	18.06	16.92	Kim et al., 2014
	S. cerevisiae	Glucose	0.408	n.d.	5.31	Romagnoli et al., 2015
Phenylalanine	*E. coli*	Glucose	62.47	1,301.46	48.12	Ding et al., 2016
para-Coumarate	*S. cerevisiae*	Glucose	1.93	26.81	7.91	Rodriguez et al., 2015
Resveratrol	*S. cerevisiae*	Glucose + ethanol	0.53	5.21	0.25	Li et al., 2015
Styrene	*E. coli*	Glucose	0.26	8.97	5.6	McKenna and Nielsen, 2011
	S. cerevisiae	Glucose	0.0.29	0.604	0.464	McKenna et al., 2014
Tyrosine	*E. coli*	Glucose	0.18	4.29	15.67	Olson et al., 2007
	E. coli	Glucose	9.7	440.91	17.76	Lütke-Eversloh and Stephanopoulos, 2007
	E. coli	Glucose	3	111.11	11.49	Chávez-Béjar et al., 2008
	E. coli	Glucose	13.8	383.33	20.9	Santos et al., 2012
Reticuline	*E. coli*	Glycerol	0.046	0.45	0.316	Nakagawa et al., 2011
	S. cerevisiae	Glucose	0.0000806	0.00083958	0.00055	DeLoache et al., 2015
	S. cerevisiae	Glucose	0.000192	0.002	0.0025	Trenchard et al., 2015

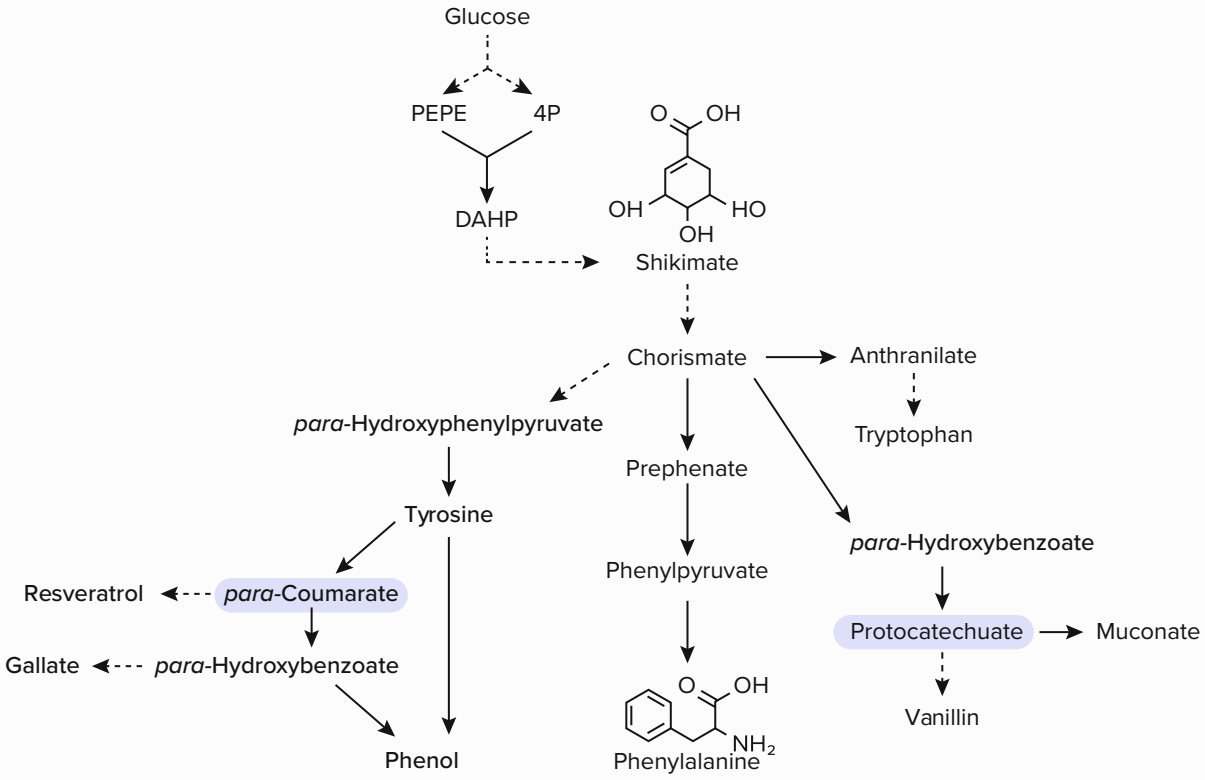
More telling are the commercial examples of bio-based production of compounds derived through branches of the shikimate pathway. Industry-led R&D investment has the potential to unlock more opportunities along this pathway.

Outcomes of Industry-Led R&D on Shikimate Pathway Derivatives

Consider coumarate and protocatechuate (PCA), which are two key intermediates and branch points of the shikimate pathway that can unlock a wide array of valuable downstream molecules. PCA, for example, is a precursor to molecules like vanillin, gallate, and catechol, which are all one to two enzymatic step conversions from PCA. Similarly coumarate is a precursor to molecules such as caffeate, ferulic acid, and resveratrol, which are produced from coumarate within a few enzymatic steps.

Recent and continued work conducted by Ginkgo Bioworks demonstrates commercially compelling flux toward these precursors, with clear room for further improvement towards the maximum biological theoretical yield. Through iterative partner-driven R&D, Ginkgo has achieved improvements in flux that match or exceed best-in-class reports in the literature and open the door for biomanufacturing a variety of downstream targets for commercial use (Liu et al, 2019, Kogure et al 2021).

The Shikimate Pathway and Branches off of Coumarate and Protocatechuate

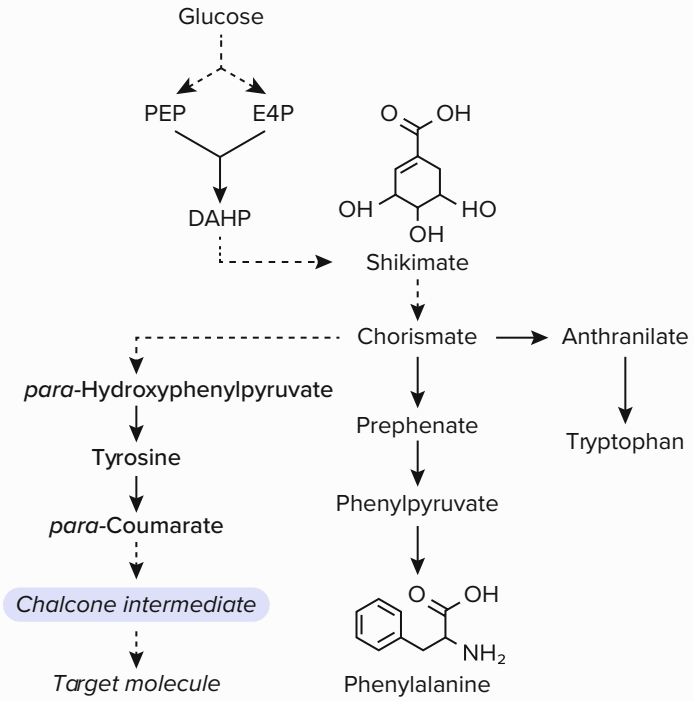


Example 1
Aromatic Molecules for Nutrition and Wellness

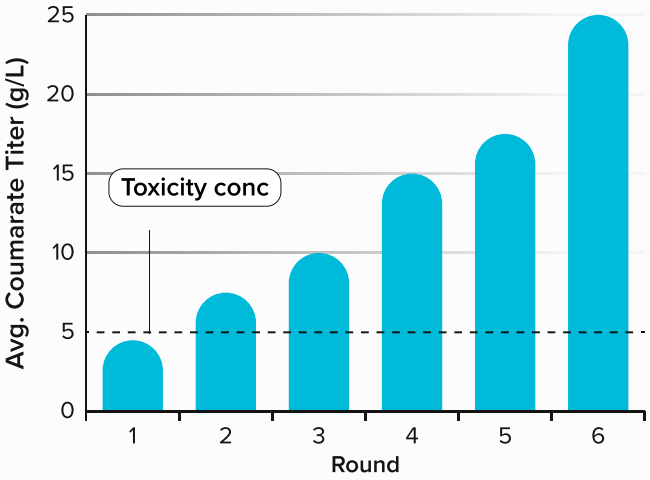
Ginkgo Bioworks is a synthetic biology R&D platform, designing and licensing custom organisms to partners who aim to commercialize biological products through fermentation, including small molecule ingredients. In one such engagement, a partner identified an aromatic target molecule that has promising nutrition and wellness applications but costly plant extraction routes and no known chemical synthesis route. Achieving a bio-based production route enables this partner to open up a larger market for the otherwise supply-limited product.

For this program, Ginkgo Bioworks narrowed down a few potential pathways branching from coumarate that could yield the target molecule in yeast. The naturally occurring enzymes along these pathways, including the pathway to precursor coumarate, were known to have poor activity and specificity, thus limiting carbon flux toward the target molecule. Ginkgo's first development objective was to create a coumarate chassis strain that exhibits significant flux to coumarate. The second development objective was to establish significant flux to a second key intermediate – a chalcone, the next known precursor to the target molecule in plants and in yeast. From there, Ginkgo achieved production of the final target molecule by yeast fermentation, which is the first known instance of bioproduction of the target molecule from glucose. The final objective of this program is to engineer the full pathway to achieve the most economically viable unit cost per kilogram of the target molecule.

The Shikimate Pathway Branches to the Partner's Target Molecule



Coumarate Titer Improvement After Six Rounds of Engineering



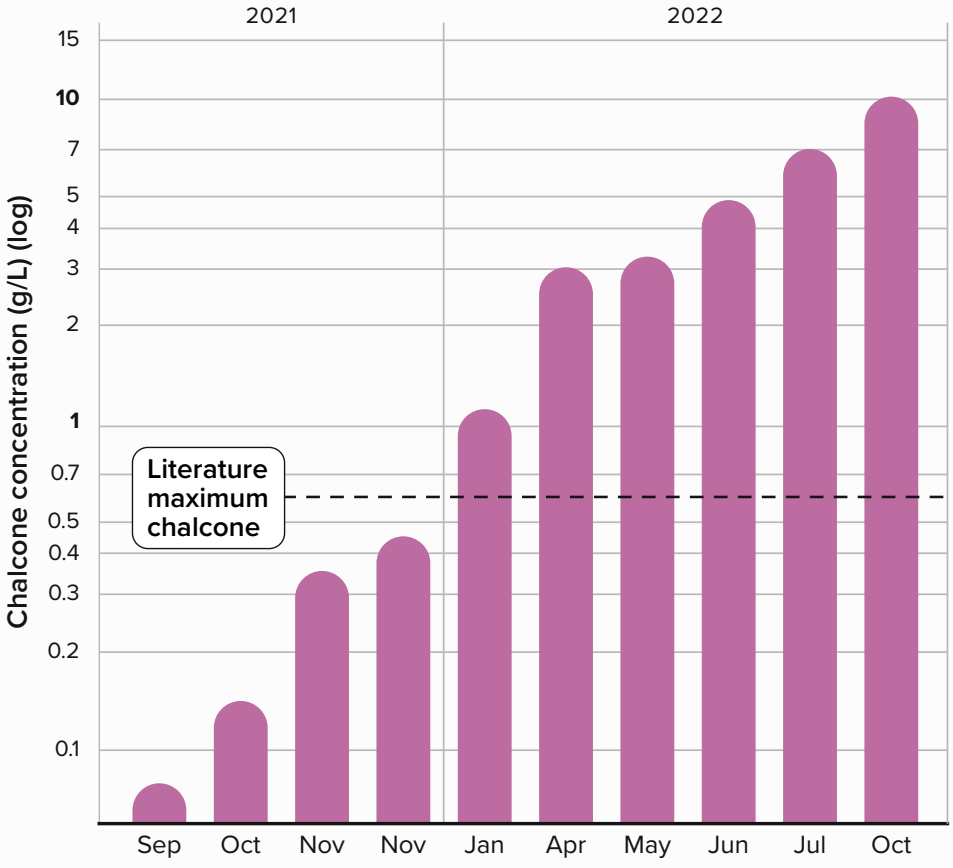
Toward the first development objective, Ginkgo engineered a *S. cerevisiae* chassis strain to produce 25 g/L of coumarate, a titer that was nearly 2× higher than previously reported in literature (Liu et al, 2019). The program achieved this milestone in 6 months, by first focusing on the pathway steps known to be hampered by poor enzymatic activity and specificity. By leveraging proprietary metagenomic databases, Ginkgo identified enzymes from other species likely to have improved activity and specificity along the path to coumarate. Subsequently, libraries were constructed to explore the impact of improved enzyme specificity and activity along the path to shikimate, as well as knockdowns

and deletions of competing pathways upstream of coumarate. Examples of knockdown and deletion targets include the glycolytic pathway downstream of PEP and the mandelate pathway downstream of chorismate.

This achievement paved the way for the next development objective: increasing flux through an intermediate chalcone. The path from coumarate to the intermediate chalcone in yeast is reportedly hampered by chalcone synthase (CHS), an enzyme with low abundance and specificity (Tong et al, 2021). Furthermore, CHS requires a central metabolite, malonyl-coenzyme A (malonyl-CoA) which is transiently expressed during yeast growth and reproductive cycles. To debottleneck this critical step, Ginkgo leveraged protein engineering to design a library of mutant CHS enzymes. In parallel, the team explored genetic designs that could yield higher pools of malonyl-CoA. In 8 months, the team achieved nearly 10 g/L production of this intermediate chalcone, 16× higher than previously reported in literature (Ref., 2020, citation provided upon request).

For the final objective of this program, Ginkgo Bioworks identified three potential pathways toward the target molecule in yeast, eventually narrowing down to one with confirmed flux from the chalcone to the target molecule. To achieve this, Ginkgo's team validated the required enzymatic steps to go from the chalcone to the target molecule by testing metagenomic enzyme libraries on the intermediate substrates. Top performing enzymes from these libraries underwent structural analysis to inform Ginkgo's protein engineering strategy towards improved specificity along the pathway. All together, the R&D program drives towards economic viability, based on our partner's target unit cost.

Intermediate Chalcone Titer Improvement After Engineering

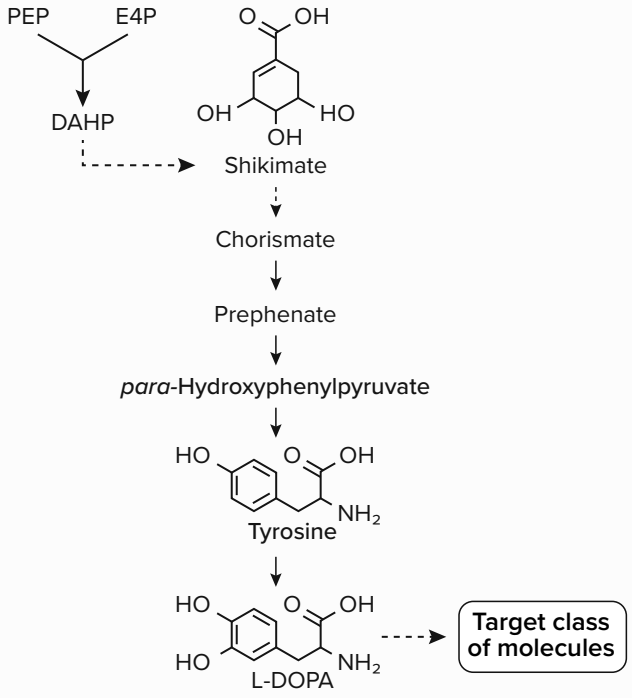


Example 2
Aromatic Molecules for Food Colorants

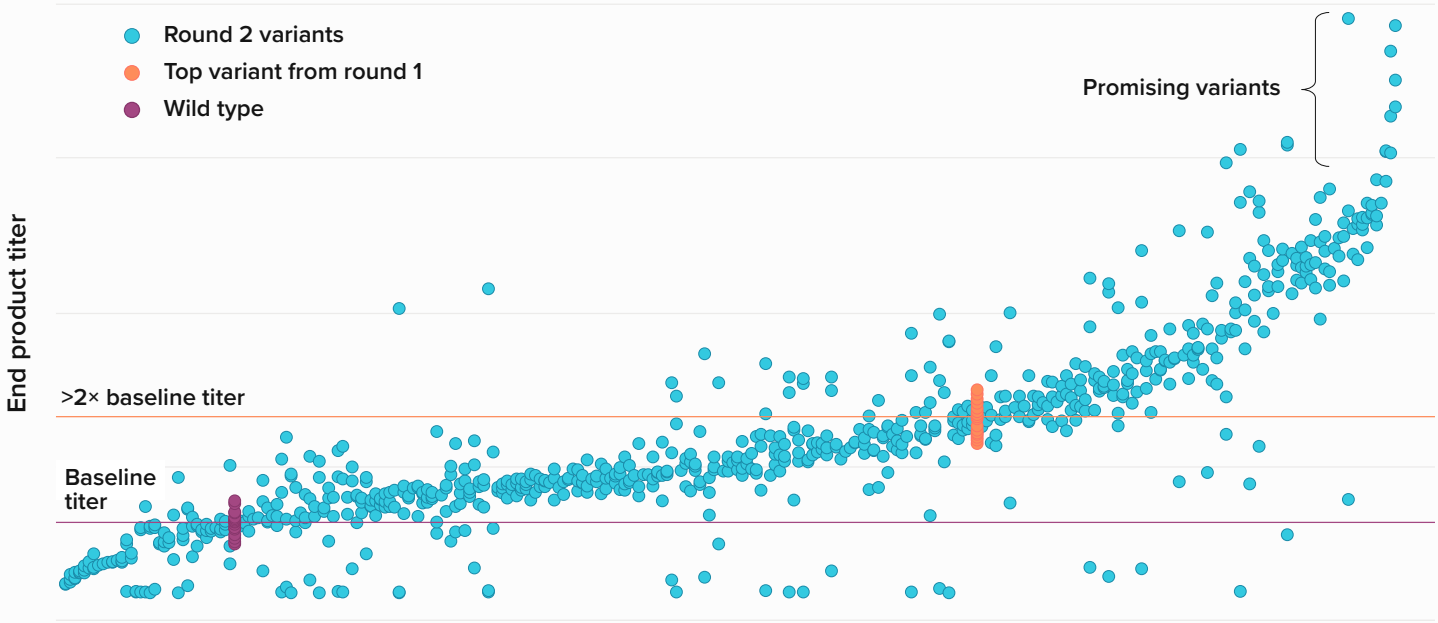
Coumarate is one enzymatic step downstream from tyrosine and two enzymatic steps downstream from phenylalanine in *S. cerevisiae*. The coumarate chassis strain – engineered to produce high titers of coumarate – provides an advanced starting point for production of tyrosine derivatives that can be used as food colorants. Ginkgo Bioworks used the coumarate chassis strain as a starting point for another partner involved in commercializing microbially produced alternative ingredients for food manufacturers.

Because the coumarate chassis strain described above had significant carbon flux through the shikimate pathway and further through to chorismate, the Ginkgo team shaved 4-6 months off engineering a strain that could support high flux toward the target class of molecules. While the final goal of this ongoing program is to optimize production of these target food colorants, the first aim is to alleviate a key bottlenecking step. In two rounds of protein engineering, Ginkgo has identified variants of a native enzyme that allow for more than four-fold improvement in the production titer of these target food colorants.

The Shikimate Pathway Branches to the Partner's Target Molecule



~300 Enzyme Variants Screened in Round 2



Example 3
Aromatic Molecules for Biobased Materials Intermediates

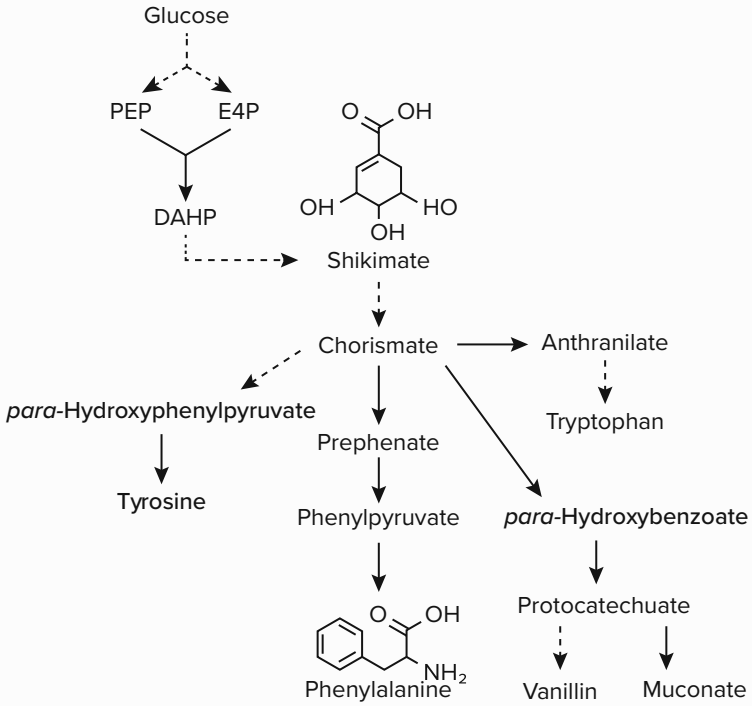
Industrial biotech R&D has yielded headstarts for bio-based production of compounds in the materials sector. In this example, Ginkgo is undertaking a program to engineer an organism that produces a third class of target molecules that will be used in the materials sector. While previous examples have focused on branch points from coumarate, this program utilizes another branch of the shikimate pathway downstream of protocatechuate (PCA). A *S. cerevisiae* chassis strain similarly engineered to produce 75 g/L of PCA served as the starting point for this program with our biomaterials partner.

Through its large scale, Ginkgo Bioworks is pursuing multiple engineering approaches in parallel to arrive at a final strain with the partner's target characteristics. In one approach, Ginkgo is evaluating metagenomic libraries for each enzyme along the pathway and selecting those that demonstrate improved specificity. Through a pooled screening approach, various combinations of specificity and gene copy numbers of the top-performing enzymes along the pathway have been tested to tune expression/abundance of the enzymes in the cell. This approach made apparent toxicity and acid tolerance issues with one of the target molecules.

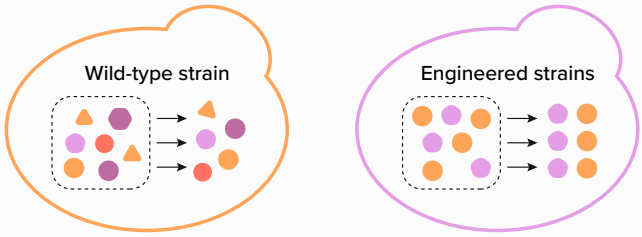
Ginkgo turned to another technique to address this issue – adaptive laboratory evolution (ALE), where proprietary technology automates evolution and applies selective pressure on microbial strains, thereby drifting them towards more tolerant phenotypes. In this instance, instead of running ALE experiments, Ginkgo could leverage prior ALE results to rationally engineer tolerance into the new strain based on known mutations in the *S. cerevisiae* strains used in these earlier experiments. Those strains can subsequently be used in this project to address toxicity, due to the commonality of global toxicity challenges in this chassis organism.

In each of these examples, existing knowledge and biological assets from prior work at Ginkgo Bioworks enabled us to achieve milestones for our partners sooner than would be possible otherwise. This process of iterative learning across projects can drive R&D cycle time down and enable lower unit costs for partners, who gain access to engineered organisms that have a higher chance of achieving theoretical biological maximum yields.

The Shikimate Pathway Branch to PCA



Desired Outcome of Strain Engineering



Choosing a Target Molecule

Before embarking on an R&D program to develop bio-based molecules, companies, and investors should consider the economic factors at play: the size and pricing structure of the end-market, projected R&D costs to achieve a strain with target performance, capital expenditure to manufacture the product at scale, and time to market to determine how quickly they can recoup their investment. Since each biological molecule and pathway is unique, with its own bottlenecks and processing requirements, understanding whether or not a specific target molecule can be manufactured through fermentation economically is an important factor to consider in any development effort. Techno-economic analysis (TEA) combines these factors to enable data-driven investment decisions.

Both the upfront R&D costs and the eventual unit costs in an operating business play an important role in determining a project's overall economic viability. Choosing attractive markets, both in size and unit margins, can help reduce the payback period. To illustrate the point, a molecule produced with a COGS of $500/kg in a 100 MT market with 50% margins would earn the same annual operating margin to pay back upfront R&D expenses as a bioproduced molecule with a COGS of $10/kg in a 45,000 MT market with 10% unit margins.

In choosing a target molecule, investors may also compare unit costs of production from microbial fermentation processes to incumbent processes (e.g. plant extraction or petroleum) by modeling the theoretical yield of each process from their respective feedstocks. For example, if maximum theoretical yield of a target molecule from *S. cerevisiae* is determined to be 0.5 g product / g substrate and the cost of the substrate is $0.6 / kg substrate, the absolute minimum cost of a product is the cost of the substrate divided by the maximum theoretical yield: $1.20 / kg of product in this example. This contributes to the unit cost floor of the target molecule that can be achieved through fermentation, and other processing inputs will contribute additional unit costs.

Indeed, some biological molecules may not have a direct COGS advantage when compared head-on to their incumbent counterparts. This could be due to fully depreciated legacy assets, tenured processes, economical raw inputs, etc. However, aspects for consideration beyond cost can include: increase in performance, supply chain resilience, preparedness for regulatory changes, and sustainability advantages. For example, a bio-based molecule could have increased and differentiated activity compared to its petroleum-derived analog, thereby enabling a formulation chemist to require less of the substance in an overall formula, thus driving down formula level costs.

Evaluating R&D Investment and Outsourcing Options

Outsourcing decisions can impact the net present value of R&D investment in a few ways. Most obviously, fixed costs associated with building internal R&D capabilities are partially replaced by variable costs of outsourcing on an as-needed basis. Outsourcing can also accelerate time-to-market. As demonstrated above, working with a partner who has an advanced understanding of the target pathway or molecule – and even an advanced capability, such as an engineered strain with significant flux through an intermediate – can trim R&D time and spend. While it remains challenging to predictably engineer biology, an R&D partner with expertise in a broad array of tools can claim a higher likelihood of achieving and potentially exceeding titer, productivity, and/or yield goals.

Ginkgo Bioworks can help partners evaluate economic viability for molecules under consideration for bioproduction. By leveraging Ginkgo's de-risked pathways, chassis strains, and world class approach to engineering biology, partners can reduce the cost of R&D, the time to market, and can improve the overall COGS of many biobased compounds. This whitepaper focused on Ginkgo's work along the shikimate pathway, one of many pathways the company has significantly de-risked.


GINKGO
BIOWORKS