An Automated Scientist to Design and Optimize Microbial Strains for the Industrial Production of Small Molecules

Amoolya H. Singh[†], Benjamin B. Kaufmann-Malaga[†], Joshua A. Lerman[†], Daniel P. Dougherty, Yang Zhang, Alexander L. Kilbo, Erin H. Wilson, Chiam Yu Ng, Onur Erbilgin, Kate A. Curran, Christopher D. Reeves, John E. Hung, Simone Mantovani, Zachary A. King, Marites J. Ayson, Judith R. Denery, Chia-Wei Lu, Phillip Norton, Carol Tran, Darren M. Platt, Joel R. Cherry, Sunil S. Chandran, Adam L. Meadows[*]

Amyris Inc., 5885 Hollis St. Ste 100, Emeryville CA 94608
[†]These authors contributed equally to this work.
*Correspondence to: meadows@amyris.com

Abstract

Engineering microbes to synthesize molecules of societal value has historically been a time consuming and artisanal process, with the synthesis of each new non-native molecule typically warranting its own separate publication. Because most microbial strain engineering efforts leverage a finite number of common metabolic engineering design tactics, we reasoned that automating these design steps would help create a pipeline that can quickly, cheaply, and reliably generate so-called microbial factories. In this work we describe the design and implementation of a computational system, an Automated Scientist we call Lila, which handles all metabolic engineering design and optimization through the design-build-test-learn (DBTL) paradigm. Lila generates metabolic routes, identifies relevant genetic elements for perturbation, and specifies the design and re-design of microbial strains in a matter of seconds to minutes. Strains specified by Lila are then built and subsequently phenotyped as part of a largely automated in-house pipeline. Humans remain in-the-loop to curate choices made by the system, helping for example to refine the metabolic model or suggest custom protein modifications. Lila attempted to build strains that could produce 454 biochemically diverse molecules with precursors located broadly throughout the metabolism of two microbial hosts, *Saccharomyces cerevisiae* and *Escherichia coli*. Notably, we observed the highest published titers for the molecule naringenin, the metabolic precursor to flavonoids. In total we created hundreds of thousands of microbial strains capable of overproducing 242 molecules, of which 180 are not native to *S. cerevisiae* or *E. coli*.

Keywords: Industrial biotechnology; cell factory; computational biology; metabolic engineering; high-throughput; biofoundry

Introduction

Many public and private organizations commonly use industrial fermentation of engineered microbes to make commercially viable or societally valuable products. These products, either a pure molecule or a mixture of molecules, include pharmaceutical ingredients[1], vitamins[2–5], flavors & fragrances[4,6], pesticide alternatives[7] , plastics[8,9] and polymers[10], lubricants[11], and dozens of others that would otherwise come from petrochemical processing, endangered plant species, or an expensive or volatile supply chain. Although industrial fermentation of engineered microbes has proven impactful in that it can deliver high purity biomolecules in sustainable fashion at stable cost, the R&D process of bringing a single biomolecule from idea to small scale production (milligrams) to applications testing (kilograms) to market (tons) is lengthy

(years to a decade or more) and expensive ($100-$200M), a barrier that prevents many new biomolecules from going to market[12].

This R&D process involves several stages of differing costs and durations. The first step involves identifying a molecule of interest. This may last from weeks to months, with a focus on market research, estimated production and purification cost, and technical feasibility. Once a molecule has been identified, the next step is to delineate the biosynthetic pathway. This is typically done manually by a trained metabolic engineer, who must either identify candidate pathways from literature or derive *de novo* pathways using expert knowledge of what biochemical reactions have precedent. Once the biosynthetic pathway has been delineated, any required enzymes that are not present in the host species must be identified, codon-optimized for expression in the host, ordered for gene synthesis, and prepared for insertion into the host. This step typically takes an additional 1-2 weeks of literature search, phylogenetic analysis, modification (for example, to remove unwanted intracellular targeting sequences), and submission of a gene synthesis order. Once the synthesized genes have been delivered, the DNA constructs must be assembled into a pathway *in vivo*, its activity confirmed, and the target molecule detected. Depending on the number of iterations required to synthesize the product, these steps take anywhere from a few months to a year. Once a proof-of-concept strain has been created that can produce detectable amounts of the molecule in the milligram range, various facets of cellular metabolism and physiology must be optimized until the strain is capable of producing the molecule at titers in the 0.5 - 5 g/L scale. This phase may include balancing enzyme co-factors, improving product export, overcoming enzymatic bottlenecks, and improving cell health and genetic stability. This optimization step tends to be the most time-consuming phase and can vary from years to a decade or more, depending on the product[12]. The next phase is to continue the optimization while starting pilot scale fermentations to produce and purify kilogram-levels of the molecule that will allow its testing for various applications. While this testing is ongoing, strain optimization continues until a cost target is finally reached that enables the scale up of production, typically another few months to a year.

Strain optimization tends to be the rate-limiting step in the R&D process both in time and cost because a vast biological search space must be explored[12] by either rational design of a desired genotype & phenotype, or random mutagenesis followed by screening for a desired phenotype. For instance, in the rational approach to strain engineering[13], there may be four potential biosynthetic routes to the target molecule starting from the carbon source (in our case sugar). Each of those routes are frequently composed of twenty or more biochemical reactions, some of which may be native to the organism and require tuning or changes to regulation, whereas others may be non-native and require the expression of heterologous genes. Each reaction may be catalyzed by hundreds of different possible enzymes. Certain enzymes may have dozens of subunit choices, with half a dozen choices for where to truncate an enzyme for optimal expression, and dozens of potential codon optimizations for optimal translation. Other enzymes, such as P450s, may require helper proteins to become active. Each enzyme may have hundreds of loci and promoter-terminator permutations in which to insert into the host genome for optimal pathway balancing. Thus, the total design space accessible to a typical strain engineer is the combinatorial space of these factors, which must be combined carefully, as factors can be fully crossed (any engineered protein could be expressed with any validated promoter), partially crossed (any protein could be integrated at any genomic locus unless it would create a loop out), or nested (each P450 may have only a few cognate reductase candidates). Together the total design space of potential strains exceeds ~10^{24}. The classical approach to strain improvement[14], which entails genome-wide or site-directed mutagenesis followed by screening for a desirable trait, is invaluable

when the targets for optimization are not known *a priori*. In this case enumerating the number of non-silent base pair changes in a 12MB genome results in ~10^{30} possible mutants. With such a vast search space, it can take hundreds of iterations of DBTL cycles to optimize the production of a molecule from milligrams to kilograms. Assuming that each DBTL cycle takes weeks to months, scaling molecule production from milligrams to kilograms can take years.

One way to reduce the risk of the R&D process is to work on multiple similar molecules at once, expecting that a few will successfully scale. To do this cost-effectively, however, requires some level of automation so that each new molecule added to the R&D portfolio does not linearly add cost in the form of time or headcount. This cost reduction can be achieved if the multiple molecules are deliberately chosen to be biosynthetically and chemically similar[15]. A second way to reduce the risk of the R&D process is to diversify the kinds of molecules in the pipeline, expecting that a few will successfully commercialize. This is more challenging in practice, as the automation and analytical chemistry may not easily generalize to dissimilar molecules (e.g. flavonoids vs. terpenes), and the strain designs cannot be shared across molecules without additional cost in the form of time or headcount.
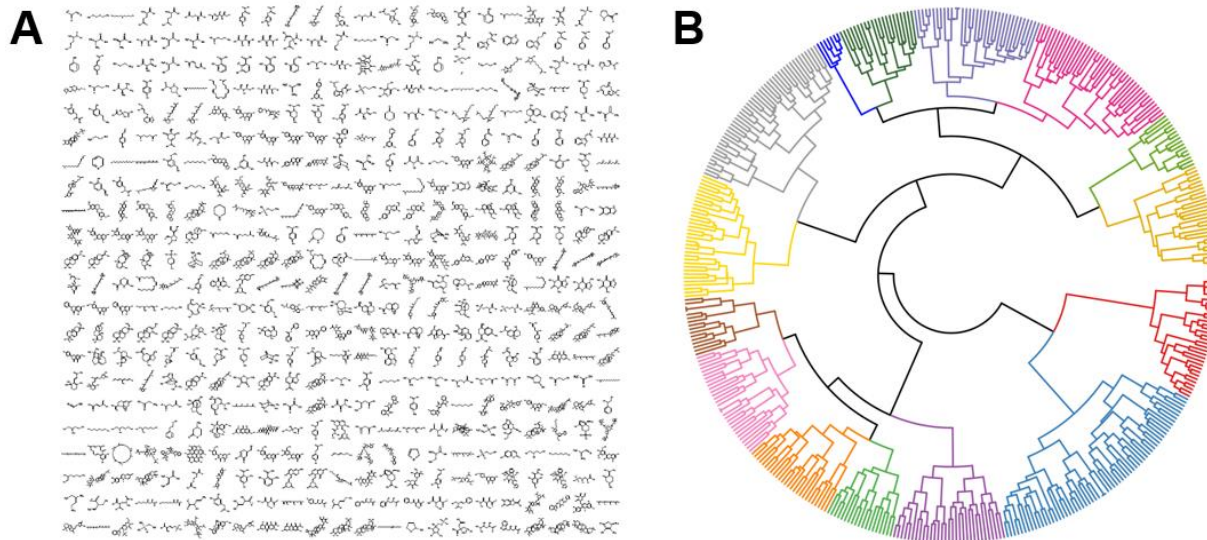

Figure 1. (A) Structures of molecules targeted by the Automated Scientist. (B) Chemical dendrogram of 454 molecules whose biosynthesis was attempted by the Automated Scientist, illustrating the structural diversity of molecules into 15 major classes. To generate this clustering, we retrieved Structure Data Files from PubChem for each molecule and used atom pair similarity (Tanimoto distance) to calculate a distance metric between molecules. We then hierarchically clustered the distance matrix of Tanimoto coefficients with Ward's linkage.

In this paper, we describe the design and implementation of an Automated Scientist software platform we named Lila that mitigates these two risks: increase the number of molecules under development, and increase the diversity of molecules under development, without significant manual intervention. Lila combines mathematical modeling and design principles generally adopted by scientists trained in the field to match a genotypic hypothesis or experimental design to the resulting phenotypic data.

The notion of automating various aspects of microbial synthesis, or even laboratory protocols, is not new. Various groups have successfully demonstrated the computational representation of biochemical reactions[16,17,18,19,20], automation of biosynthetic route finding[21,22,23], and laboratory protocols[24,25,26,27,28,29,30]. In the area of computational chemistry, the idea of automated synthesis is also gaining traction[31,32,33,34,35].

As impressive as these previous efforts have been, several factors set our work apart from previous efforts. Most important is scale. We have attempted more molecules in parallel than has previously been reported by at least an order of magnitude. We also are the first to report full industrialization, in terms of the development of high-throughput pipelines to undertake design, build, test, and learn phases with near-touchless handoffs. and rate of molecule optimization (in terms of the time elapsed between first detection of a molecule and its scale-up to kilogram amounts). With the Lila platform we have shortened the timeline for molecule proof-of-concept production (i.e., zero to detectable) from months to weeks, decreased the cost of molecule development to <$10M per molecule (10-fold reduction), and simultaneously handled several hundred molecules in a multiplexed pipeline. Over the course of 105 partially overlapping DBTL cycles, Lila made >100,000 *in silico* designs targeting the production of 454 small molecules (Figure 1), ordered 1,850 genes for synthesis selected entirely by algorithm, created 32,000 distinct microbial strains (involving laboratory operations that assembled 280 million bps of DNA and transformed ~180 million bps of DNA), and analyzed more than 10,000,000 data points. Taken together, we achieved a hit design in half of the molecules targeted and established a reproducible and largely automated process for biosynthesizing almost any small organic molecule in a microbial host.

Results

An Automated Scientist has modules akin to the cognitive functions of a human scientist

The Lila Automated Scientist software is broken into subtasks or cognitive functions that a human scientist would employ to approach a complex problem. Just as a human scientist perceives new information, learns, remembers, thinks, and executes, so too there are Lila software modules corresponding to memory, perception, reasoning, and execution. Lila was initialized with rules known to experts in the field in the Route Finding and Strain Designer modules, which interact with the Knowledge Store database that stores genotype and phenotype data. These data are used by a Strain Designer module analogous to the cognitive function of a human scientist that creates abstract genetic design rules, and decides among which of several strain improvement strategies (enzyme improvement, mutagenesis, rational engineering, or deconvolution of mutations) to pursue next. The abstract design rules are compiled into a detailed DNA sequence implementation by the Genotype Generator module, analogous to the execution function of a human scientist. Finally, the Phenotype Analyzer and Process Analyzer modules, which implement the perception and learning functions of a human scientist, analyze multivariate data and reason about uncertainty. All these modules work together with the rest of the Amyris pipeline to select molecules and analyze strains. These modules are related as shown in Figure 2.
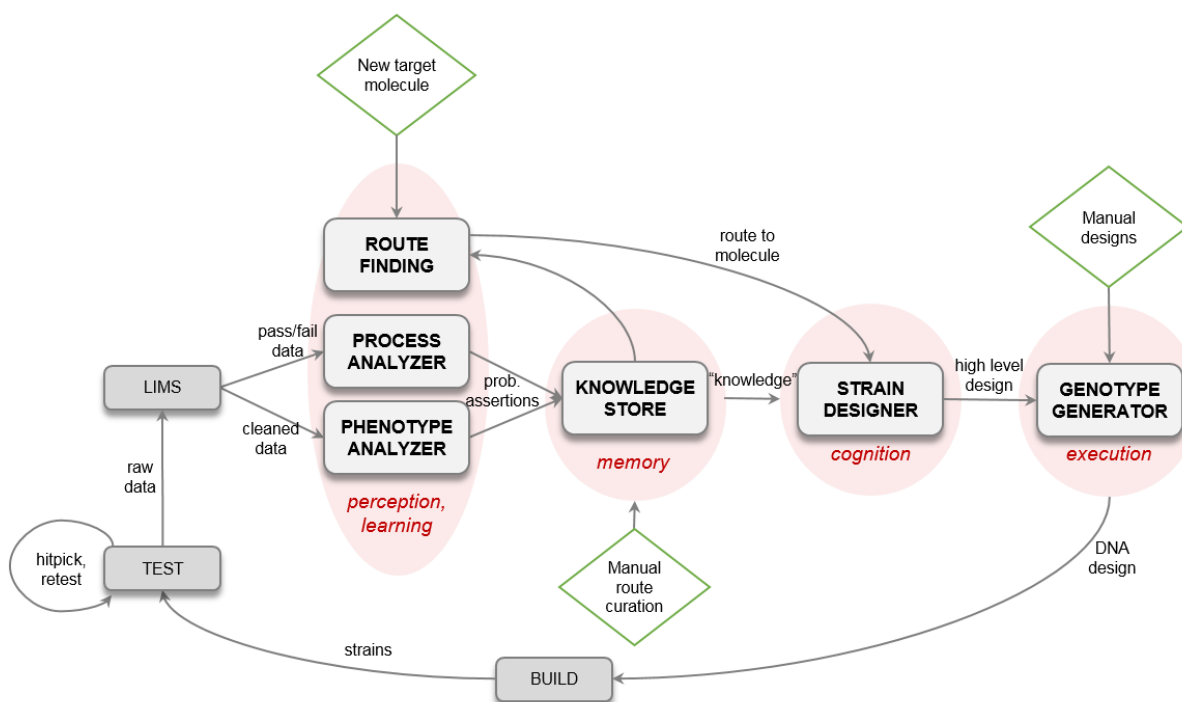
Figure 2. Parts of the Lila platform are analogous to a human scientist's workflow. Green diamonds represent input points in which human scientists interact with Lila; "prob. assertions" are probabilistic assertions, i.e. scientific hypotheses with associated probabilities

More concretely, to go from target molecule to physical strain design the Lila platform makes decisions at several levels of abstraction (Figure 3). The lowest level of abstraction (called the Molecules layer) is a set of decisions around what molecules to include within the model, and how they are represented *in silico*. For example, intracellular metabolites that are transient as a reaction proceeds may be excluded to simplify a model. Next the feasible biochemical reactions within the chosen host organism (Reactions layer) are

encoded and associated with physical properties such as their Gibbs free energy. These reactions are then aggregated into biochemical routes that connect sugar and other input nutrients to the final desired target molecule (Routes layer), which introduces a vast combinatorial challenge. The fourth step (the Design layer) involves choosing which biochemical reactions to physically engineer within the strains. Next (in the Proteins layer) the Automated Scientist software decides what specific proteins to introduce into the cell or modify (in the case of native genes). Finally, it adds regulation and genomic context to the design (Layouts layer) and fully specifies the genotype down to the base pair level. Details on each of these layers are described in the remainder of this paper.

Layer	Module(s)	Schematic	Factors	Constraints
Molecules	Knowledge Store		• Presence of InChI keys, SMILES, etc. • Inclusion of transient or protein-bound intermediates	• Toxicity to host
Reactions	Knowledge Store		• Thermodynamic favorability • Validated in biology	• Balanced (mass, charge, etc.) • Precedent in vitro or in a biological setting.
Routes	Strain Designer		• Selection of reactions • Ordering of reactions • Cofactor requirements • Theoretical yield	• High flux precursors • Number of reactions below construction limit • Limited toxic intermediates
Designs	Strain Designer		• Reactions to regulate • Native reactions to downregulate	• All non-native reactions must be included. (N.B. organism must be specified at this level.)
Proteins	Strain Designer	protA protB protC	• Codon optimization • Organism of origin • Truncations and/or localization sequences	• Cofactor requirements met • Nontoxic to host
Layouts	Strain Designer	protC protA protB protC	• Locus of integration • Regulatory elements (promoters, terminators, etc.) • Organism	• Loop-out risk in vivo • Strain assembly pipeline

Figure 3. The Lila platform makes decisions at multiple distinct levels, from the biomolecules and reactions to consider to the final genomic layout. Each of these levels has a number of factors that can be experimentally modified as well as a number of constraints that must be considered.

Lila's Algorithm Generated Routes to 454 Diverse Biochemical Targets

To establish what biochemical reactions are permissible within the chosen microbial host, we populated the Knowledge Store with over 20,000 biochemical reactions, including those from the MetaNetX database[36], into a repository we refer to as the Amyris Universal Set of Reactions (Figure 4A). We next annotated these reactions using a combination of public data and in-house curation, allowing Lila to filter challenging reactions based on their thermodynamic infeasibility, lack of experimental support in the literature, or other factors. Lila's Route Finder algorithm then stitched these reactions together into a biochemical route from glucose to the target molecule of interest (Figure 4B) and optimal routes for 454 molecules were identified for each model organism (Figure 4C).
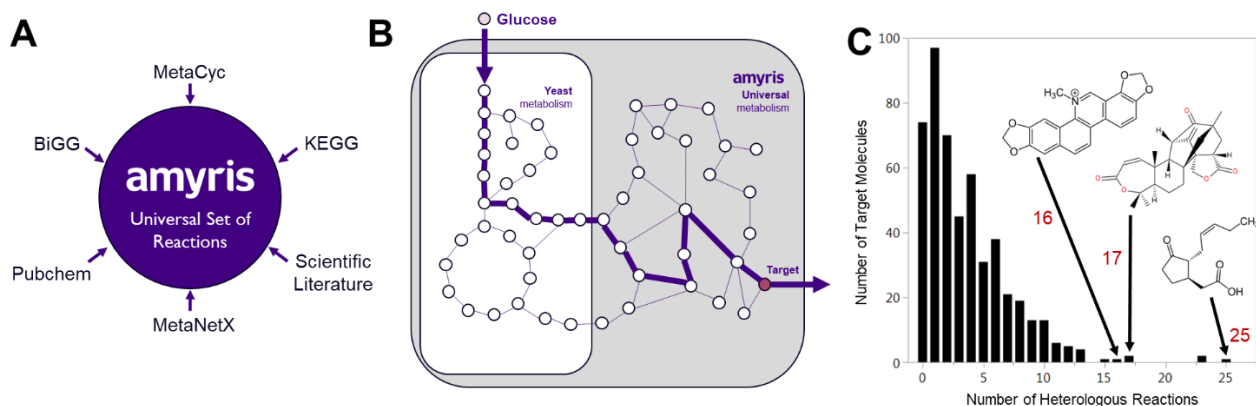
Figure 4. Automated route finding resulted in biochemical pathways going from glucose to 454 target molecules in both S. cerevisiae and E. coli. (A) Combing reactions from a variety of public databases and in house efforts resulted in a universal set of over 2.9 million compartmentalized biochemical reactions. (B) Biochemical pathways derived from this set of reactions span both native yeast metabolism (embedded white box) as well as non-native reactions (grey region). (C) The majority of pathways require 4 or fewer non-native reactions to complete the pathway, but some molecules needed 15 or more heterologous steps.

For each target molecule, Lila's Route Finder emits an ordered list of native and heterologous enzymes that catalyze successive reactions to obtain the target molecule. Most target molecules require 1-5 heterologous enzymes to catalyze the pathway (Figure 4C). For each enzyme represented as an Enzyme Commission number[37] (EC#), however, there are dozens to hundreds of enzymes encoded by various organisms that all catalyze the desired reaction. To decide which enzyme to use from which organism, we devised an algorithm that retrieves all possible amino acid sequences from Uniprot[38] corresponding to each EC#, and adaptively chooses n or fewer enzyme variants per EC#, where n is decided by an internal pipeline capacity model. Because the human procedure for selecting enzymes involves reading the literature, we implemented literature-mining algorithms that search for probabilistically enriched co-occurrences of any enzyme aliases and EC#s in all PubMed abstracts, and detect keywords such as organism names, whether the enzyme has been purified, expressed, or assayed (SI Figure 7). In addition, we constructed phylogenetic trees for each group of enzyme homologs (data not shown) and retrieved information from the BRENDA database[39] for any published information about kinetic parameters such as k_m and k_{cat}.

To design an algorithm capable of working on par with a human scientist, we interviewed a panel of Amyris scientists with a cumulative 100+ years of strain engineering experience. The questions in the interview probed what criteria they used to select pathway enzymes for testing, how they thought about codon optimization, how they identified peptide targeting sequences, etc. Table 2 of the Supplementary Information summarizes their approaches and provided a starting point for prioritizing which approaches to automate in code. In a period of 12 months, this algorithm selected 1,850 enzymes for gene synthesis, and annotated them for correct insertion into hundreds of strains making molecules of interest.

Lila's Strain Designer Explores Combinatorial Genetic Space Using Design-of-Experiments and Linear Optimization

Following the identification of a biochemical pathway and associated proteins, Lila's next challenge was to convert these high-level choices into genomic layouts (Figure 3, last row). This step involves making decisions about genomic location (e.g., what location in which to integrate each selected protein), regulation

(e.g., which promoters and terminators to include), and engineering feasibility (e.g., whether tandem or inverted repeats will be allowed, to ensure genetic stability).

To explore this vast combinatorial design space we employed a Design-of-Experiments based approach. As there are many flavors of designed experiments[40], we generated a short-list and filtered it to variants that could be practically constructed by Amyris' strain engineering pipeline and could plausibly generate metabolic pathway variants, which generally have categorical or constrained factors. The short-list includes *(i)* full factorial designs: n^k runs for k factors with n levels each, typically two levels, main effects only, *(ii)* fractional factorial designs: n^{k-p} runs, for k factors with n levels each, where p is the number of possible interactions between factors, *(iii)* definitive screening designs: *~2k* runs for k factors with up to three levels each, assessing main effects & interactions, and *(iv)* Box-Behnken designs: response surface methodology with up to three levels per factor, assessing interactions. The list of factors and levels are detailed in SI Table 1.

Finally, we tied this Design-of-Experiments framework into a genotype specification so that combinatorial designs could easily be specified without needing to spell out each design variant afresh in code for each experimental cycle. Generating genotypes requires translating the high-level designs, which largely consisted of a list of enzymes to insert, overexpress, down-regulate, or remove, into low-level DNA constructs and strain specifications for Amyris' Automated Strain Engineering Pipeline (ASE). Communicating a design specification to ASE is done via Genotype Specification Language[41] (GSL), but for the specific task of automating thousands of designs, we needed to encode repeated sets of complex GSL statements (higher-level design constructs) into simpler function calls. We therefore implemented GSL functions to express a given number of genes at a given locus, drive a given gene with a given promoter, and so on. In consultation with biologists on the design of the ratified base strains for each microbial species, we encoded several GSL functions in advance that could insert the heterologous genes into the pre-existing landing pads.

Because our pipeline has finite capacity to construct strains and we were pursuing many targets at once, our samplings of these combinatorial spaces were often quite sparse. As a result we frequently began with a small number of strain designs at diverse points throughout the combinatorial space, and iterated until we converged to fewer successful designs. Once an algorithm identified the proteins that should be overexpressed, we implemented a number of strategies to formalize the re-design[42], trying different homologs of the protein, and finding a new route that avoids the protein altogether (if the protein cannot be further overexpressed or if no suitable homologs can be identified).

Lila Generated Hit Strains Producing Hundreds of Non-Native Small Molecules that Span Yeast Metabolism

Once strains were constructed using the Strain Designer module and measured for product titer, it was possible to begin the strain optimization phase. To identify and correct sub-optimalities in a biosynthetic pathway within a live strain, several factors and corresponding measurements were considered. First, the amount of flux passing through each metabolic node or reaction in the wild-type strain guides the type and degree of genetic or process-level perturbation needed. The expression of pathway enzymes is also important, and one wants to avoid scenarios in which pathway enzymes have poor expression while competing steps have plentiful expression. Third, the pool size of a given metabolite matters: the higher the substrate concentration, the higher the thermodynamic driving force towards the target molecule of interest. Thermodynamics (e.g., reference ΔG^{m}) plays a role particularly if there is an off-pathway step that is undesirable, but thermodynamically very favorable. Finally, while build-ups of pathway intermediates may in some cases represent wasted carbon not going to the desired product, and therefore undesirable, they may also be necessary in some cases to drive thermodynamically unfavorable reactions within a given pathway.

Each of these issues requires a different treatment in terms of the diagnostic approach. To close the loop between data analysis and strain re-design, we wrote a dozen algorithms that take an inductive approach in which data, rather than rules supplied by human domain experts, drive the design effort. We also relied on biologists to examine the strain designs being emitted by Lila and suggest changes or improvements. Phenotypic data such as next-generation sequencing, optical densities, titers, and different flavors of proteomics and metabolomics are mapped by these algorithms to extensively curated metabolic models derived from published models for *S. cerevisiae*[43] and *E. coli*[44]. Conceptually, this work introduces a form of ensemble learning whereby multiple weak learning algorithms (i.e., simple algorithms with substantial error rates) are run in parallel to yield an overall lower error rate.

The results of this pipeline are shown in Figure 5 for the target molecule naringenin, a key platform node. Strains designed by Lila (Figure 5B) produced naringenin in a microtiter plate-based model at a wide range of output titers, from zero to >1.5 g/L (Figure 5A). To our knowledge, this titer of naringenin exceeds the highest published titer in the literature[45–48].
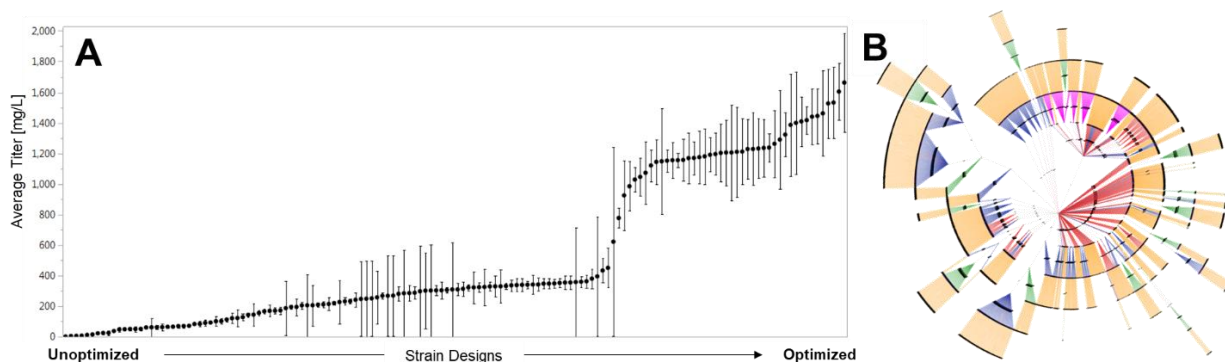
Figure 5. Successive strain and design improvements to produce the target molecule naringenin. (A) Titers in microtiter plates. Each dot represents the mean of all titer measurements for a given strain, and the error bars represent the standard error over N≥3 replicates. Only strains with N≥3 titer measurements are shown (n=163). (B) Comprehensive strain lineage map for naringenin strains (n=2,879) in a circular layout, with the common base strain in the center. Two nodes are linked by a directed edge color-coded based on the type of genetic modification used to modify the parent strain: promoter swap or knockout (green), integration of heterologous genes at one of three different genomic landing pad positions (blue, red, and pink), and plasmid curing (orange).

The overall results of tasking Lila to develop strains producing 454 different molecules are shown in Figure 6. Over the course of 105 partially overlapping DBTL cycles, 242 distinct molecules were detected at levels above the host organism's native level, 140 of which achieved titers over 10 mg/L in microtiter plates, while 96 of which achieved titers over 100 mg/L. A subset of strains for 77 distinct molecules were tested in a 2L fed-batch bioreactors[49] to further interrogate strain potential and generate material for applications testing. This often led to an order of magnitude increase in observed titer (Figure 6).

A comprehensive description of all attempted molecules, their relative molecular complexity[50,51], and two measures of pathway complexity, can be found in the SI Table 3. We observe that more complex molecules and pathways generally resulted in strains with lower titers (SI Figure 8, Figure 9, and Figure 10). For example, the average molecule complexity at titer tiers above the >0.1 g/L threshold were all significantly lower than the average molecule complexity for molecules that were not produced at levels above the host organism's native levels. However, for detection of <u>any</u> production that was above the host organism's native level of synthesis, there was not a statistically significant effect of molecule complexity. This implies that molecule complexity may be an important factor in the automated optimization of strains producing those molecules by our pipeline, but it does not significantly impact the low level "proof-of-concept" production of those molecules. The strongest correlation with titer tier was the minimum number of heterologous enzymes required for a molecule's synthesis. This is not surprising given that for each heterologous enzyme in the pathway introduced into the host organism, there is a probability that either the protein or required reaction chemistry fail to function in the new host, making the probability of successful production drop with each additional heterologous gene required. Despite that, our methods produced multiple hits for pathways containing 8 to 11 heterologous genes.
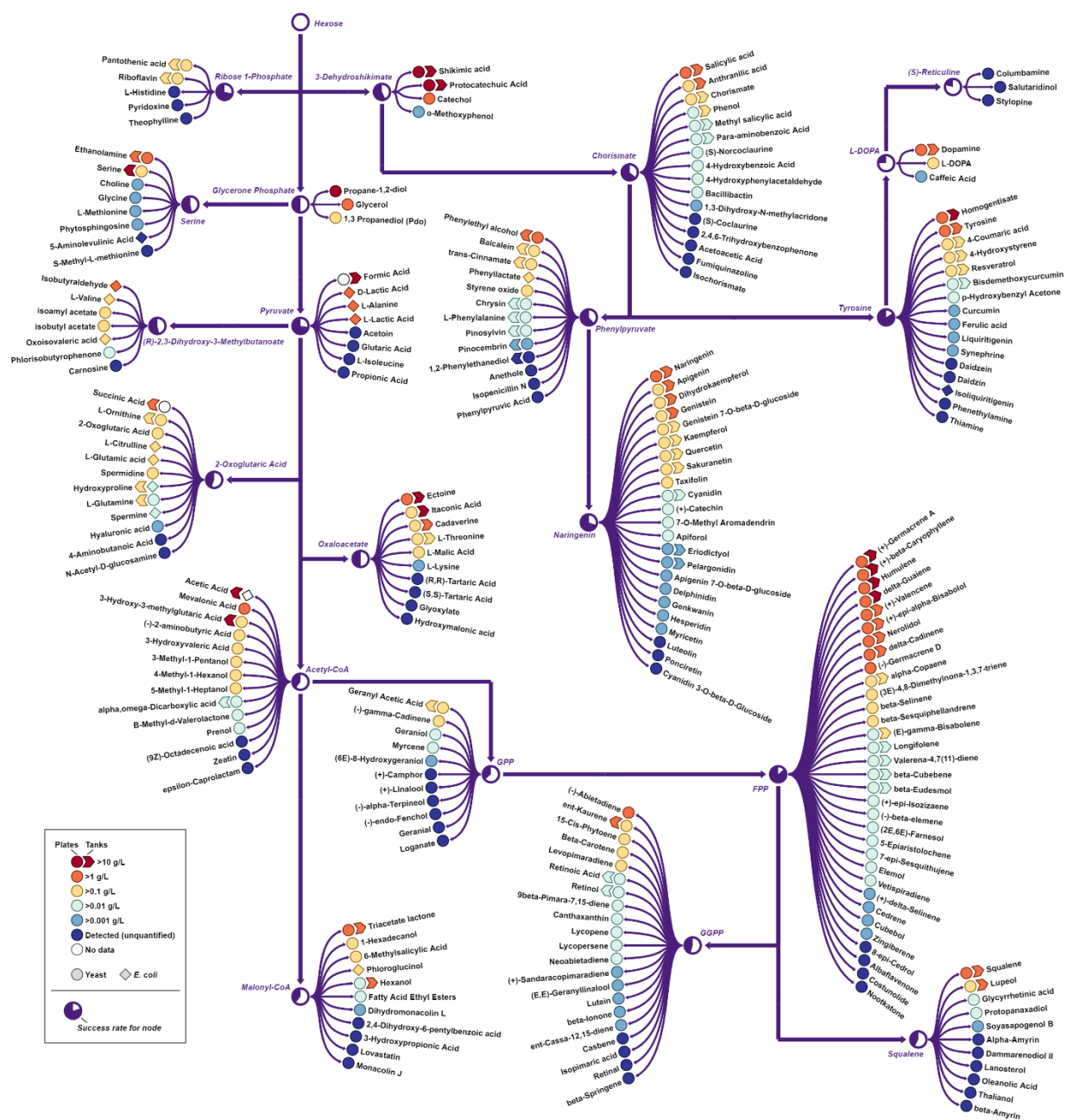
Figure 6. Titer results across metabolism show titer hits in 21 of 22 different metabolic nodes. *A simplified metabolic map shows all targets with detectable titers. Success rates for metabolic nodes are the fraction of all targeted metabolites for which the molecule was detected in either 96-well microtiter plates containing 40 g/L sucrose[52] or 2L fed-batch fermentation tanks[49]. Molecule titers were quantified by generic GC-MS or LC-MS/MS methods[51,52]. The map was generated using the Escher[55] software package.*

11

Methods

Route Finder Algorithm

The Route Finder is based on flux-balance analysis (FBA)[56]. Starting with published genome-scale metabolic models for *S. cerevisiae*[43] and *E. coli*[44], we extended these models to include all known heterologous reactions using our Universal Set of Reactions. Furthermore, all reversible reaction stoichiometries were split into two unidirectional reactions such that all reaction fluxes **v** were strictly non-negative. We then ran FBA on this extended model **S** with an objective function weight vector **w. w** was defined with negative weights for all reactions. Reactions that were native to the model organism received a small negative weight and heterologous reactions were assigned a ~100x larger negative weight. This effectively penalized the use of any heterologous reactions unless required to produce the target metabolite. The Route Finder performed FBA on these extended models with COBRApy v0.3.2 to find the minimal set of heterologous reactions that could produce an arbitrarily small demand flux c on the target metabolite for overproduction:

$$maximize \ \boldsymbol{w}^T \boldsymbol{v}$$

$$such \ that \ \mathbf{Sv} = \mathbf{0}, \mathbf{0} \leq \mathbf{v} \leq \mathbf{v_{max}}, and \ v_{demand}^{target} \geq c$$

The resulting flux distributions from these models were further analyzed for thermodynamic feasibility, overall stoichiometry, and maximum theoretical yield. Roughly a quarter of the routes were checked manually by Amyris strain engineers, which fed back into successive refinements of Lila's Route Finder algorithm by altering elements of **w** or **v**$_{max}$.

Enzyme Selection Algorithm

The enzyme selection algorithm filters amino acid sequences using scientific criteria prioritized in the following order: *(i)* the fraction of PubMed abstracts mentioning an enzyme of interest and all its aliases, that also mention the target molecule of interest or a specific host organism, *(ii)* existence of k_m values in BRENDA, *(iii)* experimental evidence for the existence of the genetic sequence at the protein or mRNA level from the Uniprot database, *(iv)* phylogenetic tree topology, and *(v)* sequence diversity from a multiple sequence alignment.

If the first pass of the algorithm chooses more variants than can be reasonably synthesized, it attempts to reduce the candidates by making the selection criteria more stringent. If there are too few variants, it attempts to add back candidates by relaxing the selection criteria. It does this iteratively for each enzyme until convergence. In many cases the algorithm starts with over 100 candidate enzymes for a given EC#, and can narrow it down to four candidates within ten iterations (<5s runtime).

The amino acid sequences selected for each EC# are then codon-optimized for expression in the desired host, most frequently at Amyris the *S. cerevisiae* strain CEN.PK2[57] using Genotype Specification Language[41] (GSL) compiler with proprietary codon optimization extensions, and stored in the Knowledge Store. To track each gene and anonymize the submission to the gene synthesis vendor, we devised a

random/unique hash key for each DNA sequence that tracks the molecule ID, Uniprot ID, EC#, source organism, and versions/parameters of algorithms that were used to generate the DNA sequences.

To decide on these criteria, we conducted roughly a dozen hours of interviews with Amyris strain engineers and binned their responses into a ranked-choice scheme summarized in SI Table 2.

Towards a Curated Knowledge Store of Molecules, Genes, and Proteins

The goal of building the Knowledge Store was to be able to store various public and in-house facts pertaining to the DBTL cycle, in order to drive automated strain design and re-design. At the outset of this effort, we conducted a gap analysis to assess what data were already routinely stored at Amyris and what additional data were needed to support Lila. Briefly, the following systems were already in place:

- **Strain** database, storing strain genotypes, a record of integrations/deletions/modifications, and all strain lineages in the history of Amyris, together with QC data

- **Part** database, storing the inventory of all DNA parts with their sequence information, and QC data provided by next-generation sequencing

- **High-throughput Screening** database, storing results of dozens of distinct 96-well plate assays in high- and medium-throughput format, together with experimental conditions

- **Fermentation & Analytical Chemistry** databases, storing results of all benchtop & full-scale fermentations, including time series traces of off-gas measurements, analytical chemistry measurements from GC/LC, mass balance calculations, and experimental conditions

- **Omics** databases, storing results of all transcriptomics, metabolomics, and proteomics experiments, along with experimental conditions

- **Sequencing** database, storing results of all whole-genome sequences and analysis therein, such as identification of SNPs, indels, transpositions, and copy number changes

As this was already quite a substantial list of databases covering Build & Test and we were keen not to reinvent the wheel, we focused on building out specific storage to support Design & Analyze (re-design). These include:

- **Molecule** database, recording public & private identifiers for every molecule & metabolite that may ever appear in a biochemical pathway or a chemical synthesis; together with information on molecule name aliases, safety, hazard/precaution, toxicity to human and microbe, and measured & predicted physico-chemical attributes such as boiling point, Kovats retention index, viscosity, partition coefficient

- **Metabolic model** database, storing all publicly known and privately curated biochemical reactions, stoichiometries of participating metabolites, subcellular localization of each metabolite, and mappings between reactions and enzymes

- **Route** database, storing lists of reactions comprising a biochemical pathway to a target molecule, allowing the storage of routes generated by different computational algorithms (flux-balance analysis, shortest-path, manually curated), and affiliated data such as maximum theoretical yield, yield-productivity cost curves, and Gibbs' free energies of reactions

13

- **Protein/enzyme/gene/organism** (central dogma) database, storing public identifiers for proteins, enzymes (EC#s), genes, and organisms, starting with the Uniprot, Brenda, Metacyc, and NCBI databases and extending these to include literature-mined and privately curated entities; literature information about enzyme activity & catalytic rates, protein multimers forming an enzyme, taxonomic lineages, mappings between enzymes and proteins, and name aliases for genes, enzymes, and proteins

- **Business-specific** databases, storing in-house information on running the DBTL pipeline such as updated molecule status, contents of a genetic design, mapping between design and DNA parts, details of gene synthesis orders and codon optimization parameters, and details of machine learning algorithms

The stores above are implemented as a Postgres relational database on a dedicated load-balanced set of servers. For data that are retrieved from public sources, we have also written automated daily, weekly, or monthly scripts that synchronize our databases with newly-available public information. The frequency of these updates depends on the frequency with which the public database is updated and the volume of new data to be imported. Every row of every database table is timestamped with create & modify times, allowing us to rapidly re-generate a copy of the database from any arbitrary point in time and recreate the dataset that may have fed a particular algorithm.

There were several technical challenges in creating, populating, and cross-indexing tables in the knowledge store. First, the naming schemes for different molecules, metabolites, and reactions had to be unified. For example, "ethanol" has 4,292 different names in different databases that all indicate C_2H_5OH; "muconic acid" has three different isomers and two charge states that needed to be reconciled, and only a specific subset of these isomers/charge states actually participate in biochemical reactions. We solved this problem by assigning each molecule/metabolite a unique numeric ID, and matching this numeric ID to all possible aliases from different databases. To reconcile charge and isomeric states, we used IUPAC International Chemical Identifiers (InChI), whose three-part format allowed us to distinguish between the bond matrix, the isomer, and the charge. Second, the database schemas had to be merged between the knowledge store and Amyris' existing database of enzymes and metabolic models. Merging the two database schemas involved merging & cross-indexing metabolites with molecules, merging enzymes, and merging reactions. Finally, we also cross-indexed Amyris's repository of hundreds of thousands of DNA parts with the Knowledge Store's schema for genes and enzymes. This was a complex effort involving lookups between Amyris' internal store for gene names, public curated databases for host organisms such as Saccharomyces Genome Database and EcoCyc, and public sequence databases such as Uniref and NCBI RefSeq. Completing this cross-indexing has enabled fast in-house searches of DNA for heterologous enzymes, as well as the ability to rapidly map novel designs back to existing Amyris strains.

Determining Factor Levels in Combinatorial Designs

We first finalized the list of factors that would be part of the combinatorial design (Table 1). We then worked against the allocated pipeline capacity as well as the base strain design to dial down the number of factors and levels so as to obtain roughly four strains/variants per molecule and to maximize re-use of DNA parts in subsequent cycles. Many factors that we hoped would float freely, such as terminator, were eventually bound by the restrictions of DNA assembly or re-use. Nevertheless, our designs cross promoter, terminator, and gene with locus in several flavors of designed experiments to aid future mapping between genotype and phenotype.

Table 1. Factors and levels used in combinatorial strain design.

Factor	Levels
Promoter	At least four promoters, up to 7
Terminator	At least 2 terminators, up to 20
Locus, and relative position within locus	4 experimentally characterized intergenic loci, and up to 4 relative positions within locus
Gene variant (organism)	# levels will depend on each gene (typically 1-4)
Codon optimization	At least two variants, up to 3
N-terminal end (outcome is dependent on each specific gene)	Not truncated, Truncated, +GB1 tag
Copy number	0 (native genes that are not overexpressed), 1, or up to 4
Orientation in DNA part	Forward, reverse

Strain optimization algorithms using metabolite data

Remove all reactions using or producing off-pathway metabolite. In this algorithm we leverage a curated table of reactions to identify all reactions in which the identified off-pathway metabolite is either a substrate or product. These reactions are next mapped to native proteins using our metabolic model. Finally, we created updated strain designs that replace the promoter of these native genes with one or more weak promoters.

Remove all single reactions that KO ability to produce off-pathway metabolite. In this algorithm each off-pathway metabolite is first used to create a metabolic flux model capable of generating, *in silico*, that metabolite. Then, each reaction in the original route is removed, and the metabolic flux model is re-run. If removal of a reaction causes the model to be unable to generate the off-pathway metabolite (i.e., the FBA solution fails to converge on a solution), then that reaction is deemed essential for producing the off-pathway metabolite. To eliminate the off-pathway metabolite, we then generated strain designs that knockout (or titrate the promoter to lower expression) these off-pathway genes.

Remove all pairs of reactions that KO ability to produce off-pathway metabolite. In some cases, the metabolic model is able to find more than one route to a given off-pathway metabolite, and in these cases no single knockout is sufficient to eliminate accumulation of that unwanted molecule. In these cases, the re-design algorithm identifies pairs of proteins/reactions that when removed together should eliminate production of the unwanted metabolite.

Knock out all reactions that parsimoniously link route to off-pathway metabolite. Finally, we recognized that many off-pathway metabolites would be generated by carbon leaking one or two reaction steps from the core pathway, for example when a native enzyme effectively diverts flux away from a step in the engineered pathway. In these cases, we ran our metabolic model to make a route to our off-pathway metabolite using a variant of Flux Balance Analysis in which all reactions in the original design were given at no cost to the objective function. This approach biased flux towards the original pathway and therefore identified possible carbon drains that branched from the original solution. These branches were then selectively downregulated.

Strain designs implementing these algorithms were generated, had DNA parts assembled, passed through a transformation pipeline, and phenotyped. As in the case with proteomics, we implemented a diverse panel of re-design algorithms. We expect the success of each algorithm to be context-dependent, for example varying depending on the off-pathway pool size, the subsystem of metabolism being engineered, or the percentage of theoretical yield of the parent strain, and future machine learning efforts can be focused on classifying when each algorithm is most likely to succeed.

Discussion

In this paper, we describe the design an implementation of a computational system to automate the entire process of microbial strain engineering and to rapidly scale the biosynthesis of molecules from milligram to kilogram titers with minimal human input. This Automated Scientist we call Lila is implemented as software modules whose functions mimic the cognitive functions of a human scientist: memory, perception, learning, and decision-making. By evolving the Automated Scientist from an expert system[58] to an empirical learning system, we were able to bootstrap its initial runs without much initial data, and use an iterative software development cycle to add features and refine models and decision-making frameworks as data were gathered.

In its maiden run as an expert system, Lila distilled its design principles from domain experts (i.e., biologists) and articulated them as computational rules. Certain principles were hard coded. For example, we taught Lila that all P450s must be paired with an appropriate cytochrome P450 reductase, and that all non-native metabolic reactions must be catalyzed by a heterologous protein. These rules were generated deductively from an underlying theory of how cellular metabolism is believed to work, and as a result did not depend in any explicit fashion on in-house data such as metabolomics, proteomics, or titer. Although a rules-based approach was exceptionally fruitful at the initial stages, leading to the creation of strains capable of making several hundred different target molecules, it is largely unable to self-correct or learn over time. As a result, the rate of strain improvement would slow and eventually stop unless the Automated Scientist can observe how strain designs fail and automatically take corrective action.

As we collected more data, the Lila software's ability to diagnose design sub-optimalities and incorporate feedback from domain experts steadily improved. This improvement was evidenced by the 2x, 4x, or 8x improvements in titer over a matter of months after Lila's maiden run, and the ability of the Designer module to rapidly learn bottlenecks and new design rules based on the decoration of its metabolic models with new mass spectrometry data.

Synthetic biology has established itself as an emerging paradigm for how drug discovery, chemical manufacturing, and elucidation of biological mechanisms will be conducted in the future. To this end, we present a framework of how computational design, machine learning, and human intelligence can be integrated to accelerate strain design and optimization to produce a variety of chemical targets via fermentation. Traditional pathway design and optimization require inputs from multiple sources, and manual intervention at multiple points in the process, thereby increasing the timeline and cost of the entire endeavor. The goal of the Lila software is to accelerate this process and provide a computational framework to remove the strain optimization bottlenecks that scientists encounter in this field. The next step towards achieving this objective is to routinely apply machine learning algorithms to make ever more sophisticated predictions on how genetic changes impact cellular physiology using data generated from all previous experiments.

17

Competing interests

All authors are currently, or were previously, Amyris stockholders.

References

1. Westfall, P. J. *et al.* Production of amorphadiene in yeast, and its conversion to dihydroartemisinic acid, precursor to the antimalarial agent artemisinin. *Proc. Natl. Acad. Sci.* **109**, E111–E118 (2012).

2. Becker, J. & Wittmann, C. Bio-based production of chemicals, materials and fuels – Corynebacterium glutamicum as versatile cell factory. *Curr. Opin. Biotechnol.* **23**, 631–640 (2012).

3. Burgess, C., O'Connell-Motherway, M., Sybesma, W., Hugenholtz, J. & van Sinderen, D. Riboflavin Production in Lactococcus lactis: Potential for In Situ Production of Vitamin-Enriched Foods. *Appl. Environ. Microbiol.* **70**, 5769–5777 (2004).

4. Piwowarek, K., Lipińska, E., Hać-Szymańczuk, E., Kieliszek, M. & Ścibisz, I. Propionibacterium spp.—source of propionic acid, vitamin B12, and other metabolites important for the industry. *Appl. Microbiol. Biotechnol.* **102**, 515–538 (2018).

5. Fang, H., Kang, J. & Zhang, D. Microbial production of vitamin B12: a review and future perspectives. *Microb. Cell Factories* **16**, (2017).

6. Carroll, A. L., Desai, S. H. & Atsumi, S. Microbial production of scent and flavor compounds. *Curr. Opin. Biotechnol.* **37**, 8–15 (2016).

7. Quiroz, R. D. la C. *et al.* Challenges and opportunities of the bio-pesticides production by solid-state fermentation: filamentous fungi as a model. *Crit. Rev. Biotechnol.* **35**, 326–333 (2015).

8. Xiao-Ran, J., Jin, Y., Xiangbin, C. & Guo-Qiang, C. Halomonas and Pathway Engineering for Bioplastics Production. in *Methods in Enzymology* vol. 608 309–328 (Elsevier, 2018).

9. Madison, L. L. & Huisman, G. W. Metabolic engineering of poly (3-hydroxyalkanoates): from DNA to plastic. *Microbiol. Mol. Biol. Rev.* **63**, 21–53 (1999).

10. Borodina, I. & Nielsen, J. Advances in metabolic engineering of yeast *Saccharomyces cerevisiae* for production of chemicals. *Biotechnol. J.* **9**, 609–620 (2014).

11. Balakrishnan, M. *et al.* Novel pathways for fuels and lubricants from biomass optimized using life-cycle greenhouse gas assessment. *Proc. Natl. Acad. Sci.* **112**, 7645–7649 (2015).

19

12.   Nielsen, J. & Keasling, J. D. Engineering Cellular Metabolism. *Cell* **164**, 1185–1197 (2016).

13.   Prather, K. L. J. & Martin, C. H. De novo biosynthetic pathways: rational design of microbial chemical factories. *Curr. Opin. Biotechnol.* **19**, 468–474 (2008).

14.   Barbour, L., Hanna, M. & Xiao, W. Mutagenesis. *Methods Mol. Biol. Clifton NJ* **313**, 121–127 (2006).

15.   Benjamin, K. R., Hill, P. W., Meadows, A. L., Singh, A. H. & Cherry, J. R. Use Cost Models to Guide R&D. *Chemical Engineering Progress* vol. SBE Supplement-Commercializing Industrial Biotechnology 44–50 (2016).

16.   Andersen, J. L., Flamm, C., Merkle, D. & Stadler, P. F. Chemical Transformation Motifs-Modelling Pathways as Integer Hyperflows. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **16**, 510–523 (2019).

17.   Andersen, J. L., Flamm, C., Merkle, D. & Stadler, P. F. Inferring chemical reaction patterns using rule composition in graph grammars. *J. Syst. Chem.* **4**, 4 (2013).

18.   Andersen, J. L., Flamm, C., Merkle, D. & Stadler, P. Rule composition in graph transformation models of chemical reactions. in (2018).

19.   Duigou, T., du Lac, M., Carbonell, P. & Faulon, J.-L. RetroRules: a database of reaction rules for engineering biology. *Nucleic Acids Res.* **47**, D1229–D1235 (2019).

20.   Gómez-Bombarelli, R. *et al.* Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules. *ACS Cent. Sci.* **4**, 268–276 (2018).

21.   Hatzimanikatis, V. *et al.* Exploring the diversity of complex metabolic networks. *Bioinforma. Oxf. Engl.* **21**, 1603–1609 (2005).

22.   Carbonell, P., Planson, A.-G., Fichera, D. & Faulon, J.-L. A retrosynthetic biology approach to metabolic pathway design for therapeutic production. *BMC Syst. Biol.* **5**, 122 (2011).

23.   Sabzevari, M., Szedmak, S., Penttilä, M., Jouhten, P. & Rousu, J. Strain design optimization using reinforcement learning. *PLOS Comput. Biol.* **18**, e1010177 (2022).

24.   Keller, B., Miller, A., Newman, G., Vrana, J. & Klavins, E. Aquarium: The Laboratory Operating System version 2.6.0. *10.5281/zenodo.2583232* (2019).

25.   Autoprotocol. https://autoprotocol.org/.

26.   Lilly Life Sciences Studio. https://investor.lilly.com/news-releases/news-release-details/eli-lilly-and-company-collaboration-strateos-inc-launch-remote.

27.   Miles, B. & Lee, P. L. Achieving Reproducibility and Closed-Loop Automation in Biological Experimentation with an IoT-Enabled Lab of the Future. *SLAS Technol. Transl. Life Sci. Innov.* **23**, 432–439 (2018).

28.   King, R. D. *et al.* The Automation of Science. *Science* **324**, 85–89 (2009).

29.   Tenhaef, N., Stella, R., Frunzke, J. & Noack, S. Automated Rational Strain Construction Based on High-Throughput Conjugation. *ACS Synth. Biol.* **10**, 589–599 (2021).

30.   Carbonell, P. *et al.* An automated Design-Build-Test-Learn pipeline for enhanced microbial production of fine chemicals. *Commun. Biol.* **1**, 66 (2018).

31.   Coley, C. W. *et al.* A robotic platform for flow synthesis of organic compounds informed by AI planning. *Science* **365**, (2019).

32.   Collins, N. *et al.* Fully Automated Chemical Synthesis: Toward the Universal Synthesizer. *Org. Process Res. Dev.* (2020) doi:10.1021/acs.oprd.0c00143.

33.   Kotsias, P.-C. *et al.* Direct steering of de novo molecular generation with descriptor conditional recurrent neural networks. *Nat. Mach. Intell.* **2**, 254–265 (2020).

34.   Wang, Y. *et al.* Acoustic Droplet Ejection Enabled Automated Reaction Scouting. *ACS Cent. Sci.* **5**, 451–457 (2019).

35.   SynFini. https://www.sri.com/case-studies/synfini/.

36.   Moretti, S. *et al.* MetaNetX/MNXref – reconciliation of metabolites and biochemical reactions to bring together genome-scale metabolic networks. *Nucleic Acids Res.* **44**, D523 (2016).

37.   *Enzyme Nomenclature: Recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology and the Nomenclature and Classification of Enzymes*. (Academic Press, 1992).

38.   UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* **47**, D506–D515 (2019).

39.   Schomburg, I. *et al.* The BRENDA enzyme information system–From a database to an expert system. *J. Biotechnol.* **261**, 194–206 (2017).

40.   Montgomery, D. C. *Design and Analysis of Experiments*. (Wiley, 2012).

41.   Wilson, E. H. *et al.* Genotype Specification Language. *ACS Synth. Biol.* **5**, 471–478 (2016).

42.   Winkler, J. D., Halweg-Edwards, A. L. & Gill, R. T. The LASER database: Formalizing design rules for metabolic engineering. *Metab. Eng. Commun.* **2**, 30–38 (2015).

43.   Mo, M. L., Palsson, B. Ø. & Herrgård, M. J. Connecting extracellular metabolomic measurements to intracellular flux states in yeast. *BMC Syst. Biol.* **3**, 37 (2009).

44.   Orth, J. D. *et al.* A comprehensive genome-scale reconstruction of Escherichia coli metabolism-- 2011. *Mol. Syst. Biol.* **7**, 535 (2011).

45.   Gao, S. *et al.* Efficient Biosynthesis of (2S)-Naringenin from p-Coumaric Acid in Saccharomyces cerevisiae. *J. Agric. Food Chem.* **68**, 1015–1021 (2020).

46.   Koopman, F. *et al.* De novo production of the flavonoid naringenin in engineered Saccharomyces cerevisiae. *Microb. Cell Factories* **11**, 155 (2012).

47.   Xu, P., Ranganathan, S., Fowler, Z. L., Maranas, C. D. & Koffas, M. A. G. Genome-scale metabolic network modeling results in minimal interventions that cooperatively force carbon flux towards malonyl-CoA. *Metab. Eng.* **13**, 578–587 (2011).

48.   Wu, J., Du, G., Zhou, J. & Chen, J. Systems metabolic engineering of microorganisms to achieve large-scale production of flavonoid scaffolds. *J. Biotechnol.* **188**, 72–80 (2014).

49.   Meadows, A. L. *et al.* Rewriting yeast central carbon metabolism for industrial isoprenoid production. *Nature* **537**, 694–697 (2016).

50.   Bertz, S. H. The first general index of molecular complexity. *J. Am. Chem. Soc.* **103**, 3599–3601 (1981).

51.   Hendrickson, J. B., Huang, P. & Toczko, A. G. Molecular complexity: a simplified formula adapted to individual atoms. *J. Chem. Inf. Comput. Sci.* **27**, 63–67 (1987).

52.   Sandoval, C. M. *et al.* Use of pantothenate as a metabolic switch increases the genetic stability of farnesene producing Saccharomyces cerevisiae. *Metab. Eng.* **25**, 215–226 (2014).

53.   Kind, T. & Fiehn, O. Advances in structure elucidation of small molecules using mass spectrometry. *Bioanal. Rev.* **2**, 23–60 (2010).

54.   Wei, Z. & Tolstikov, V. Pattern Recognition and Pathway Analysis with Genetic Algorithms in Mass Spectrometry Based Metabolomics. *Algorithms* **2**, (2009).

55.   King, Z. A. *et al.* Escher: A Web Application for Building, Sharing, and Embedding Data-Rich Visualizations of Biological Pathways. *PLoS Comput. Biol.* **11**, e1004321 (2015).

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

57.   Salazar, A. N. *et al.* Nanopore sequencing enables near-complete de novo assembly of Saccharomyces cerevisiae reference strain CEN.PK113-7D. *FEMS Yeast Res.* **17**, (2017).

58.   Peter Jackson. *Introduction to Expert Systems*. (Addison-Wesley Longman, Incorporated, 1998).

Supplementary Information

Table 2. Ranked-choice scoring of number of strain engineers who would take each of the enzyme selection approaches as their first, second, third, or fourth choice. A -1 indicates that it is a strategy that they would explicitly avoid at first/second/third/etc.

Rank	Approach	1	2	3	4	Score
1	Identify previously successful enzymes from literature	5				20
2	Screen phylogenetic diversity	2	3			17
3	Prioritize proximity to target host	2	1			11
5	Truncate organelle targeting sequences	1			4	8
5	Try multiple codon variants		2	1		8
5	Assay enzyme in vivo		2	1		8
7	Consult kinetic data		2			6
8	Identify known mutants with improved activity		1	1		5
10	Look for transporters	1				4
10	Look for splice isoforms	1				4
10	Improve enzyme			2		4
12	Titrate promoters	-1	1			-3
14	Use enzymes with iron-sulfur cluster	-1				-4
14	Use enzymes with modular PKS	-1				-4
14	Use P450-catalyzed reaction	-1				-4
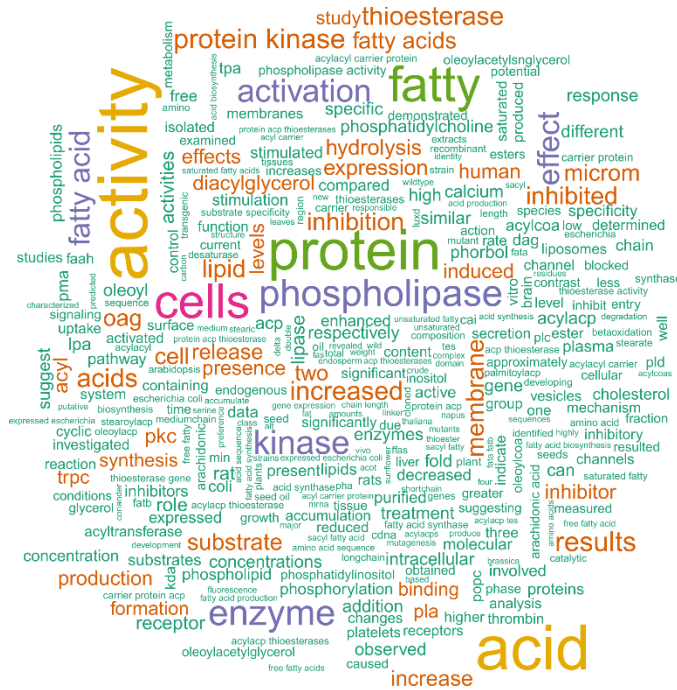
Figure 7. Word-cloud obtained from literature-mining for the enzyme "oleoyl hydrolase", EC# 3.2.1.14.

	>10 g/L	>1 g/L	>0.1 g/L	>0.01 g/L	>0.001 g/L	Detected (unquantified)	Not detected above host org…
Count	16	49	79	70	37	66	213
U95	216	275	318	392	487	396	482
Avg	169	228	269	329	394	337	434
L95	122	182	221	265	301	278	385
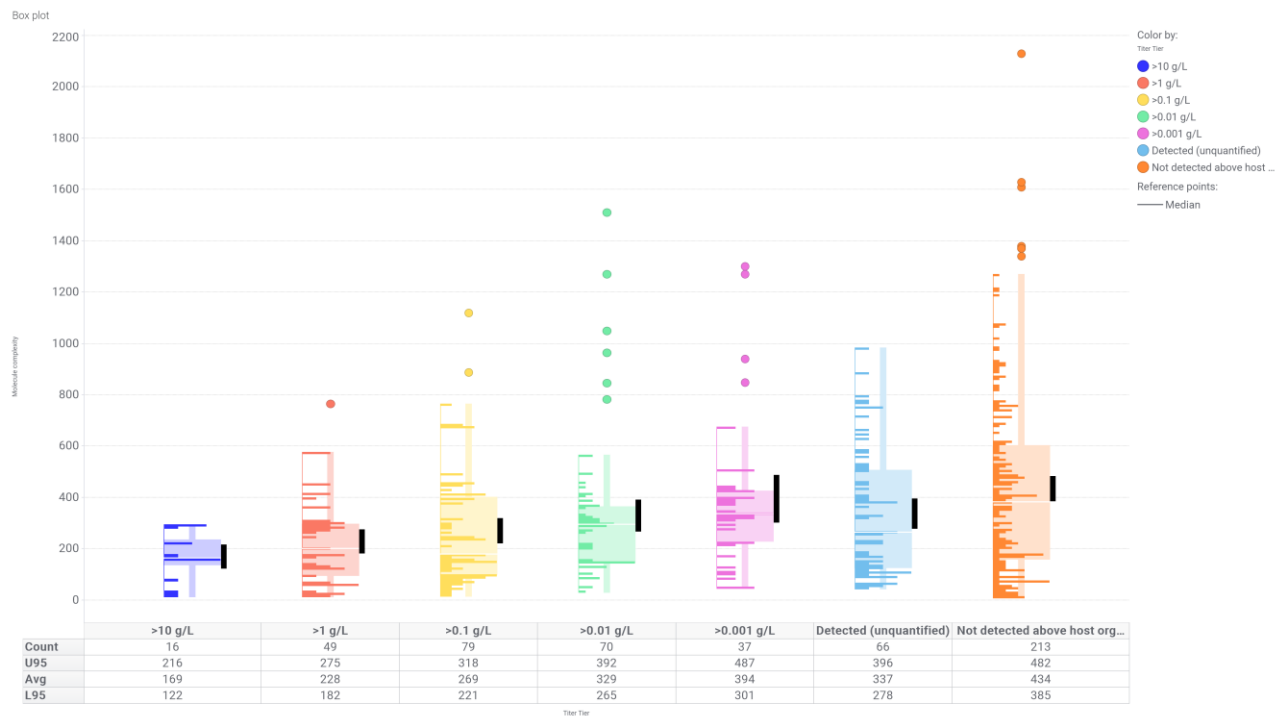
Figure 8. Box-plot of the distribution of molecule complexities at each titer tier. The black bars represent the 95% CI of the mean. The population size (Count), mean (Avg), upper 95% (U95) and lower 95% (L95) confidence interval bounds are listed explicitly in the statistics table below each plot. Data points include both 96-well plate and 2L tank observed titers as separate data points.

	>10 g/L	>1 g/L	>0.1 g/L	>0.01 g/L	>0.001 g/L	Detected (unquantified)	Not detected above host org…
Count	16	49	79	70	37	66	213
U95	1.39	1.8	2.8	2.97	4.3	3.87	5.51
Avg	0.938	1.39	2.27	2.49	3.46	3.18	4.97
L95	0.482	0.98	1.73	2	2.62	2.49	4.43
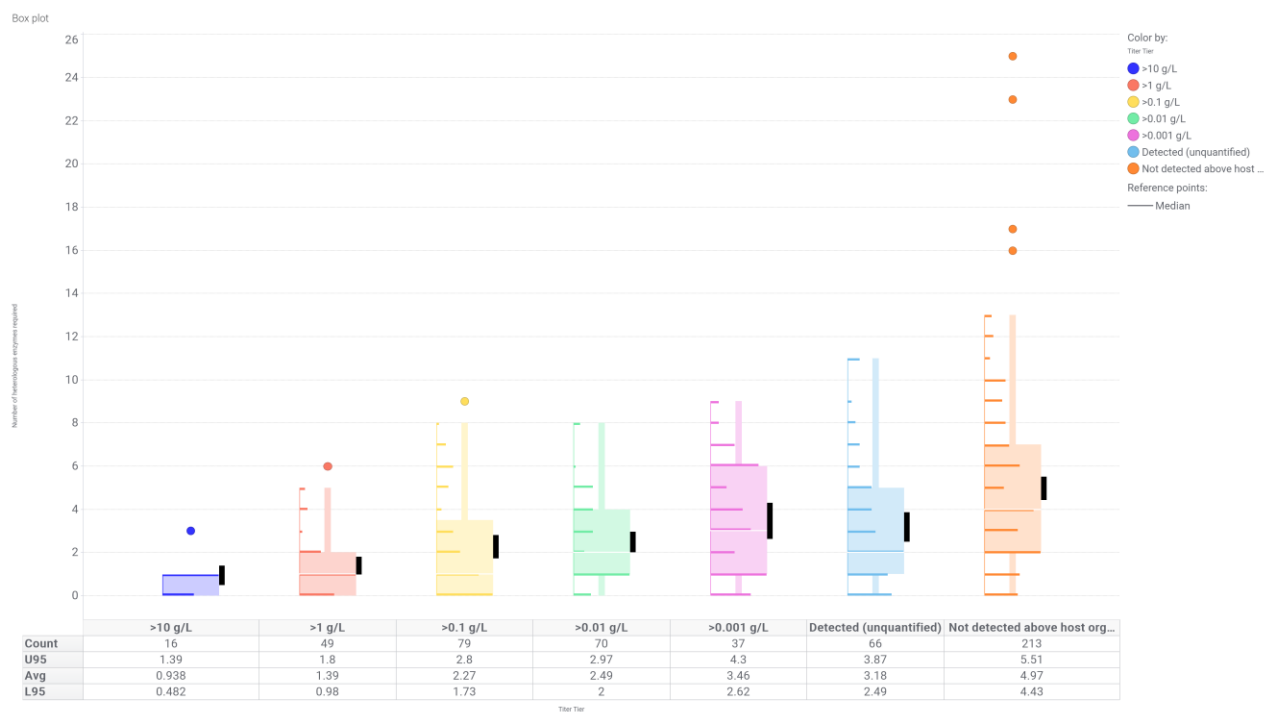
Figure 9. Box-plot of the distributions of the minimum number of heterologous genes required to synthesize the target molecules at each titer tier Note that in some cases, additional heterologous enzymes were added to improve flux above the capabilities of

25

the host pathway . The black bars represent the 95% CI of the mean. The population size (Count), mean (Avg), upper 95% (U95) and lower 95% (L95) confidence interval bounds are listed explicitly in the statistics table below each plot. Data points include both 96-well plate and 2L tank observed titers as separate data points.
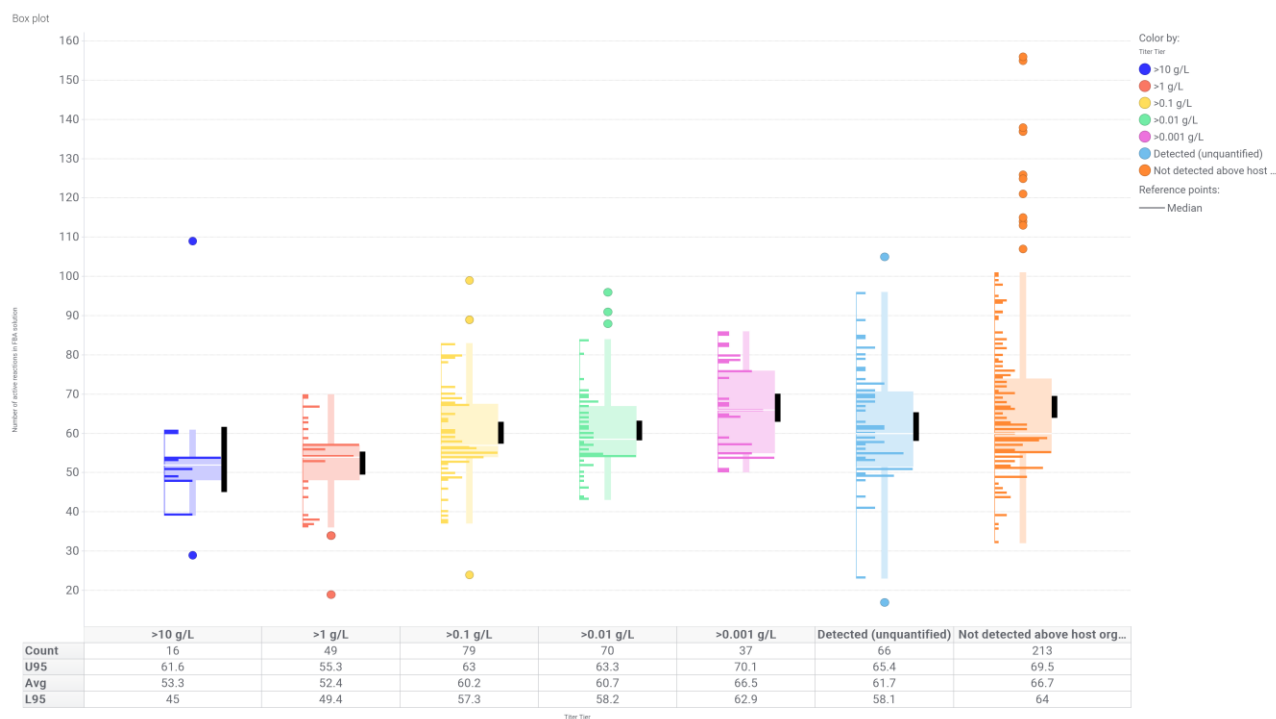
Figure 10. Box-plot of the distributions of the number of active reactions in the maximal yield FBA solution from the Route Finder for the target molecules at each titer tier. The black bars represent the 95% CI of the mean. The population size (Count), mean (Avg), upper 95% (U95) and lower 95% (L95) confidence interval bounds are listed explicitly in the statistics table below each plot. Data points include both 96-well plate and 2L tank observed titers as separate data points.

Table 3. List of all molecules attempted, the highest titer tier observed, the organism the highest titer was observed in, whether the titer was observed in a plate of fermentation tank, the molecule complexity, the number of heterologous genes required to produce the compound, and the overall number of active reactions required to achieve maximal yield in the FBA solution predicted by Lila's Route Finder module.

	molecule common name	INCHI_KEY	Host	Scale	Titer Tier	Metabolic node	Molecule complexity	Min. num. hetero-logous genes required	Num. active rxns in FBA solution
1	protocatechuic acid	YQUVCSBJEUQKSH-UHFFFAOYSA-N	Yeast	96-well plate	>10 g/L	3-Dehydroshikimate	157	1	39
2	protocatechuic acid	YQUVCSBJEUQKSH-UHFFFAOYSA-N	Yeast	2L tank	>10 g/L	3-Dehydroshikimate	157	1	39
3	Shikimic acid	JXOHGGNKMLTUBP-HSUXUTPPSA-N	Yeast	96-well plate	>10 g/L	3-Dehydroshikimate	222	0	51
4	Shikimic acid	JXOHGGNKMLTUBP-HSUXUTPPSA-N	Yeast	2L tank	>10 g/L	3-Dehydroshikimate	222	0	51
5	3-Hydroxy-3-methylglutaric acid	NPOAOTPXWNWTSH-UHFFFAOYSA-N	Yeast	2L tank	>10 g/L	Acetyl-CoA	158	1	49
6	Acetic Acid	QTBSBXVTEAMEQO-UHFFFAOYSA-N	E. coli	2L tank	>10 g/L	Acetyl-CoA	31	0	29
7	Propane-1,2-diol	DNIAPMSPPWPWGF-UHFFFAOYSA-N	Yeast	96-well plate	>10 g/L	Glycerone phosphate	21	1	48
8	L-Serine	MTCFGRXMJLQNBG-REOHCLBHSA-N	Yeast	2L tank	>10 g/L	L-Serine	73	0	53
9	Homogentisate	IGMNYECMUMZDDF-UHFFFAOYSA-N	Yeast	2L tank	>10 g/L	L-Tyrosine	168	0	48
10	Ectoine	WQXNXVUDBPYKBA-YFKPBYRVSA-N	Yeast	2L tank	>10 g/L	Oxaloacetate	177	3	61
11	Itaconic acid	LVHBHZANLOWSRM-UHFFFAOYSA-N	Yeast	2L tank	>10 g/L	Oxaloacetate	158	3	60
12	Formic Acid	BDAGIHXWWSANSR-UHFFFAOYSA-N	Yeast	2L tank	>10 g/L	Pyruvate	10	1	109
13	(+)-beta-Caryophyllene	NPNUFJAVOOONJE-IOMPXFEGSA-N	Yeast	2L tank	>10 g/L	trans,trans-Farnesyl diphosphate	293	1	54
14	(+)-Germacrene A	XMRKUJJDDKYUHV-DFSVIBJJSA-N	Yeast	2L tank	>10 g/L	trans,trans-Farnesyl diphosphate	278	1	54
15	delta-Guaiene	YHAJBLWYOIUHHM-GUTXKFCHSA-N	Yeast	2L tank	>10 g/L	trans,trans-Farnesyl diphosphate	295	1	54
16	Humulene	FAMPSKZZVDUYOS-HRGUGZIWSA-N	Yeast	2L tank	>10 g/L	trans,trans-Farnesyl diphosphate	287	1	54
17	isobutyraldehyde	AMIMRNSIRUDHCM-UHFFFAOYSA-N	E. coli	96-well plate	>1 g/L	(R)-2,3-Dihydroxy-3-methylbutanoate	31	1	38
18	Succinic Acid	KDYFGRWQOYBRFD-UHFFFAOYSA-N	Yeast	2L tank	>1 g/L	2-Oxoglutaric acid	93	0	36
19	Dopamine	VYFYYTLLBUKUHU-UHFFFAOYSA-N	Yeast	96-well plate	>1 g/L	3,4-Dihydroxy-L-phenylalanine	119	2	57
20	Dopamine	VYFYYTLLBUKUHU-UHFFFAOYSA-N	Yeast	2L tank	>1 g/L	3,4-Dihydroxy-L-phenylalanine	119	2	57
21	Catechol	YCIMNLLNPGFGHC-UHFFFAOYSA-N	Yeast	96-well plate	>1 g/L	3-Dehydroshikimate	63	2	46
22	mevalonic acid	KJTLQQUUPVSXIM-ZCFIWIBFSA-N	Yeast	96-well plate	>1 g/L	Acetyl-CoA	123	0	44
23	Anthranilic acid	RWZYAGGXGHYGMB-UHFFFAOYSA-N	Yeast	2L tank	>1 g/L	Chorismate	136	0	57
24	Salicylic acid	YGSDEFSMJLZEOE-UHFFFAOYSA-N	Yeast	96-well plate	>1 g/L	Chorismate	133	2	53
25	Salicylic acid	YGSDEFSMJLZEOE-UHFFFAOYSA-N	Yeast	2L tank	>1 g/L	Chorismate	133	2	53
26	(-)-Abietadiene	BBPXZLJCPUPNGH-CMKODMSKSA-N	Yeast	96-well plate	>1 g/L	Geranylgeranyl diphosphate	449	2	57
27	ent-Kaurene	ONVABDHFQKWOSV-ZJXASSBSSA-N	Yeast	2L tank	>1 g/L	Geranylgeranyl diphosphate	445	2	57
28	Glycerol	PEDCQBHIVMGVHV-UHFFFAOYSA-N	Yeast	96-well plate	>1 g/L	Glycerone phosphate	25	0	39
29	Ethanolamine	HZAXFHJVJLSVMW-UHFFFAOYSA-N	Yeast	96-well plate	>1 g/L	L-Serine	10	1	34
30	Ethanolamine	HZAXFHJVJLSVMW-UHFFFAOYSA-N	Yeast	2L tank	>1 g/L	L-Serine	10	1	34
31	L-Tryptophan	QIVBCDIJIAJPQS-VIFPVBQESA-N	Yeast	2L tank	>1 g/L	L-Tryptophan	245	0	69
32	Homogentisate	IGMNYECMUMZDDF-UHFFFAOYSA-N	Yeast	96-well plate	>1 g/L	L-Tyrosine	168	0	48
33	L-Tyrosine	OUYCCCASQSFEME-QMMMGPOBSA-N	Yeast	96-well plate	>1 g/L	L-Tyrosine	176	0	53
34	L-Tyrosine	OUYCCCASQSFEME-QMMMGPOBSA-N	Yeast	2L tank	>1 g/L	L-Tyrosine	176	0	53
35	hexanol	ZSIAUFGUXNUGDI-UHFFFAOYSA-N	Yeast	2L tank	>1 g/L	Malonyl-CoA	27	4	59
36	Triacetate lactone	NSYSSMYQPLSPOD-UHFFFAOYSA-N	Yeast	96-well plate	>1 g/L	Malonyl-CoA	200	1	38

No	Product	InChIKey	Host	Scale	Titer	Precursor			
37	Triacetate lactone	NSYSSMYQPLSPOD-UHFFFAOYSA-N	Yeast	2L tank	>1 g/L	Malonyl-CoA	200	1	38
38	Apigenin	KZNIFHPLKGYRTM-UHFFFAOYSA-N	Yeast	2L tank	>1 g/L	Naringenin	411	5	67
39	Dihydrokaempferol	PADQINQHPQKXNL-LSDHHAIUSA-N	Yeast	2L tank	>1 g/L	Naringenin	392	5	70
40	Genistein	TZBJGXHYKVUXJN-UHFFFAOYSA-N	Yeast	2L tank	>1 g/L	Naringenin	411	6	64
41	Naringenin	FTVWIRXFELQLPI-UHFFFAOYSA-N	Yeast	96-well plate	>1 g/L	Naringenin	363	4	67
42	Naringenin	FTVWIRXFELQLPI-UHFFFAOYSA-N	Yeast	2L tank	>1 g/L	Naringenin	363	4	67
43	Cadaverine	VHRGRCVQAFMJIZ-UHFFFAOYSA-N	Yeast	2L tank	>1 g/L	Oxaloacetate	25	1	63
44	Ectoine	WQXNXVUDBPYKBA-YFKPBYRVSA-N	Yeast	96-well plate	>1 g/L	Oxaloacetate	177	3	61
45	Phenylethyl alcohol	WRMNZCZEMHIOCP-UHFFFAOYSA-N	Yeast	96-well plate	>1 g/L	Phenylpyruvate (charge:-1)	65	0	57
46	Phenylethyl alcohol	WRMNZCZEMHIOCP-UHFFFAOYSA-N	Yeast	2L tank	>1 g/L	Phenylpyruvate (charge:-1)	65	0	57
47	D-Lactic Acid	JVTAAEKCZFNVCJ-UWTATZPHSA-N	E. coli	96-well plate	>1 g/L	Pyruvate	59	1	37
48	L-Alanine	QNAYBMKLOCPYGJ-REOHCLBHSA-N	E. coli	96-well plate	>1 g/L	Pyruvate	62	0	19
49	L-Lactic Acid	JVTAAEKCZFNVCJ-REOHCLBHSA-N	E. coli	96-well plate	>1 g/L	Pyruvate	59	1	37
50	Lupeol	MQYXUWHLBZFQQO-QGTGJCAVSA-N	Yeast	2L tank	>1 g/L	Squalene	766	2	57
51	Squalene	YYGNTYWPHWGJRM-UHFFFAOYSA-N	Yeast	96-well plate	>1 g/L	Squalene	578	0	57
52	Squalene	YYGNTYWPHWGJRM-UHFFFAOYSA-N	Yeast	2L tank	>1 g/L	Squalene	578	0	57
53	(-)-Germacrene D	GAIBLDCXCZKKJE-RXJOXMPGSA-N	Yeast	96-well plate	>1 g/L	trans,trans-Farnesyl diphosphate	266	1	54
54	(+)-beta-Caryophyllene	NPNUFJAVOOONJE-IOMPXFEGSA-N	Yeast	96-well plate	>1 g/L	trans,trans-Farnesyl diphosphate	293	1	54
55	(+)-epi-alpha-Bisabolol	RGZSQWQPBWRIAQ-GJZGRUSLSA-N	Yeast	96-well plate	>1 g/L	trans,trans-Farnesyl diphosphate	284	1	56
56	(+)-epi-alpha-Bisabolol	RGZSQWQPBWRIAQ-GJZGRUSLSA-N	Yeast	2L tank	>1 g/L	trans,trans-Farnesyl diphosphate	284	1	56
57	(+)-Germacrene A	XMRKUJJDDKYUHV-DFSVIBJJSA-N	Yeast	96-well plate	>1 g/L	trans,trans-Farnesyl diphosphate	278	1	54
58	(+)-Valencene	QEBNYNLSCGVZOH-NFAWXSAZSA-N	Yeast	96-well plate	>1 g/L	trans,trans-Farnesyl diphosphate	297	1	54
59	(+)-Valencene	QEBNYNLSCGVZOH-NFAWXSAZSA-N	Yeast	2L tank	>1 g/L	trans,trans-Farnesyl diphosphate	297	1	54
60	delta-Cadinene	FUCYIEXQVQJBKY-ZFWWWQNUSA-N	Yeast	96-well plate	>1 g/L	trans,trans-Farnesyl diphosphate	304	1	54
61	delta-Cadinene	FUCYIEXQVQJBKY-ZFWWWQNUSA-N	Yeast	2L tank	>1 g/L	trans,trans-Farnesyl diphosphate	304	1	54
62	delta-Guaiene	YHAJBLWYOIUHHM-GUTXKFCHSA-N	Yeast	96-well plate	>1 g/L	trans,trans-Farnesyl diphosphate	295	1	54
63	Humulene	FAMPSKZZVDUYOS-HRGUGZIWSA-N	Yeast	96-well plate	>1 g/L	trans,trans-Farnesyl diphosphate	287	1	54
64	Nerolidol	FQTLCLSUCSAZDY-UHFFFAOYSA-N	Yeast	96-well plate	>1 g/L	trans,trans-Farnesyl diphosphate	269	1	56
65	Nerolidol	FQTLCLSUCSAZDY-UHFFFAOYSA-N	Yeast	2L tank	>1 g/L	trans,trans-Farnesyl diphosphate	269	1	56
66	isoamyl acetate	MLFHJEHSLIIPHL-UHFFFAOYSA-N	Yeast	96-well plate	>0.1 g/L	(R)-2,3-Dihydroxy-3-methylbutanoate	87	0	59
67	isobutyl acetate	GJRQTCIYDGXPES-UHFFFAOYSA-N	Yeast	96-well plate	>0.1 g/L	(R)-2,3-Dihydroxy-3-methylbutanoate	77	0	50
68	L-Valine	KZSNJWFQEVHDMF-BYPYZUCNSA-N	E. coli	96-well plate	>0.1 g/L	(R)-2,3-Dihydroxy-3-methylbutanoate	90	0	24
69	oxoisovaleric acid	QHKABHOOEWYVLI-UHFFFAOYSA-N	E. coli	96-well plate	>0.1 g/L	(R)-2,3-Dihydroxy-3-methylbutanoate	115	0	40
70	2-Oxoglutaric acid	KPGXRSRHYNQIFN-UHFFFAOYSA-N	Yeast	96-well plate	>0.1 g/L	2-Oxoglutaric acid	171	0	49
71	Hydroxyproline	PMMYEEVYMWASQN-UHFFFAOYSA-N	E. coli	2L tank	>0.1 g/L	2-Oxoglutaric acid	125	3	52
72	L-Citrulline	RHGKLRLOHDJJDR-BYPYZUCNSA-N	E. coli	96-well plate	>0.1 g/L	2-Oxoglutaric acid	171	0	53
73	L-Glutamic acid	WHUUTDBJXJRKMK-VKHMYHEASA-N	E. coli	96-well plate	>0.1 g/L	2-Oxoglutaric acid	145	0	46
74	L-Glutamine	ZDXPYRJPNDTMRX-VKHMYHEASA-N	Yeast	2L tank	>0.1 g/L	2-Oxoglutaric acid	146	0	49
75	L-Ornithine	AHLPHDHHMVZTML-BYPYZUCNSA-N	Yeast	96-well plate	>0.1 g/L	2-Oxoglutaric acid	95	0	55
76	L-Ornithine	AHLPHDHHMVZTML-BYPYZUCNSA-N	Yeast	2L tank	>0.1 g/L	2-Oxoglutaric acid	95	0	55
77	Spermidine	ATHGHQPFGPMSJY-UHFFFAOYSA-N	Yeast	96-well plate	>0.1 g/L	2-Oxoglutaric acid	57	0	78
78	3,4-Dihydroxy-L-phenylalanine	WTDRDQBEARUVNC-LURJTMIESA-N	Yeast	96-well plate	>0.1 g/L	3,4-Dihydroxy-L-phenylalanine	209	1	53

#	Compound	InChIKey	Host	Scale	Titer	Precursor		
79	(-)-2-aminobutyric acid	QWCKQJZIFLGMSD-VKHMYHEASA-N	Yeast	96-well plate	>0.1 g/L	Acetyl-CoA	72	2 60
80	3-Hydroxy-3-methylglutaric acid	NPOAOTPXWNWTSH-UHFFFAOYSA-N	Yeast	96-well plate	>0.1 g/L	Acetyl-CoA	158	1 49
81	3-hydroxyvaleric acid	REKYPYSUBKSCAT-UHFFFAOYSA-N	Yeast	96-well plate	>0.1 g/L	Acetyl-CoA	80	4 61
82	3-methyl-1-pentanol	IWTBVKIGCDZRPL-UHFFFAOYSA-N	Yeast	96-well plate	>0.1 g/L	Acetyl-CoA	35	4 70
83	4-methyl-1-hexanol	YNPVNLWKVZZBTM-UHFFFAOYSA-N	Yeast	96-well plate	>0.1 g/L	Acetyl-CoA	44	6 67
84	5-methyl-1-heptanol	KFARNLMRENFOHE-UHFFFAOYSA-N	Yeast	96-well plate	>0.1 g/L	Acetyl-CoA	53	8 59
85	Pantothenic acid	GHOKWGTUZJEAQD-ZETCQYMHSA-N	Yeast	96-well plate	>0.1 g/L	alpha-D-Ribose 1-phosphate	239	0 80
86	Pantothenic acid	GHOKWGTUZJEAQD-ZETCQYMHSA-N	Yeast	2L tank	>0.1 g/L	alpha-D-Ribose 1-phosphate	239	0 80
87	Riboflavin	AUNGANRZJHBGPY-SCRDCRAPSA-N	Yeast	96-well plate	>0.1 g/L	alpha-D-Ribose 1-phosphate	680	0 89
88	Riboflavin	AUNGANRZJHBGPY-SCRDCRAPSA-N	Yeast	2L tank	>0.1 g/L	alpha-D-Ribose 1-phosphate	680	0 89
89	Anthranilic acid	RWZYAGGXGHYGMB-UHFFFAOYSA-N	Yeast	96-well plate	>0.1 g/L	Chorismate	136	0 57
90	Chorismate	WTFXTQVDAKGDEY-HTQZYQBOSA-N	Yeast	96-well plate	>0.1 g/L	Chorismate	392	0 54
91	Chorismate	WTFXTQVDAKGDEY-HTQZYQBOSA-N	E. coli	2L tank	>0.1 g/L	Chorismate	392	0 38
92	Phenol	ISWSIDIOOBJBQZ-UHFFFAOYSA-N	Yeast	2L tank	>0.1 g/L	Chorismate	46	1 50
93	(-)-gamma-Cadinene	WRHGORWNJGOVQY-RBSFLKMASA-N	Yeast	96-well plate	>0.1 g/L	geranyl diphosphate (charge:-3)	282	3 51
94	Geranyl acetic acid	HIGQPQRQIQDZMP-DHZHZOJOSA-N	Yeast	96-well plate	>0.1 g/L	geranyl diphosphate (charge:-3)	233	3 55
95	Geranyl acetic acid	HIGQPQRQIQDZMP-DHZHZOJOSA-N	Yeast	2L tank	>0.1 g/L	geranyl diphosphate (charge:-3)	233	3 55
96	15-Cis-Phytoene	YVLPJJIGOMTXXLP-BHLJUDRVSA-N	Yeast	96-well plate	>0.1 g/L	Geranylgeranyl diphosphate	887	1 55
97	Beta-Carotene	OENHQHLEOONYIE-JLTXGRSLSA-N	Yeast	96-well plate	>0.1 g/L	Geranylgeranyl diphosphate	1,120	5 57
98	ent-Kaurene	ONVABDHFQKWOSV-ZJXASSBSSA-N	Yeast	96-well plate	>0.1 g/L	Geranylgeranyl diphosphate	445	2 57
99	Levopimaradiene	ASPVQUYRFYUDSC-CMKODMSKSA-N	Yeast	96-well plate	>0.1 g/L	Geranylgeranyl diphosphate	449	2 55
100	1,3 Propanediol (Pdo)	YPFDHNVEDLHUCE-UHFFFAOYSA-N	Yeast	96-well plate	>0.1 g/L	Glycerone phosphate	12	2 39
101	L-Serine	MTCFGRXMJLQNBG-REOHCLBHSA-N	Yeast	96-well plate	>0.1 g/L	L-Serine	73	0 53
102	L-Serine	MTCFGRXMJLQNBG-REOHCLBHSA-N	Yeast	96-well plate	>0.1 g/L	L-Serine	73	0 53
103	Glucobrassicin	DNDNWOWHUWNBCK-JZYAIQKZSA-M	Yeast	96-well plate	>0.1 g/L	L-Tryptophan	673	9 99
104	Indigo	COHYTHOBJLSHDF-BUHFOSPRSA-N	E. coli	96-well plate	>0.1 g/L	L-Tryptophan	453	2 61
105	L-Tryptophan	QIVBCDIJIAJPQS-VIFPVBQESA-N	Yeast	96-well plate	>0.1 g/L	L-Tryptophan	245	0 69
106	4-Coumaric acid	NGSWKAQJJWESNS-ZZXKWVIFSA-N	Yeast	96-well plate	>0.1 g/L	L-Tyrosine	178	1 58
107	4-Coumaric acid	NGSWKAQJJWESNS-ZZXKWVIFSA-N	E. coli	2L tank	>0.1 g/L	L-Tyrosine	178	1 56
108	4-Hydroxystyrene	FUGYGGDSWSUORM-UHFFFAOYSA-N	Yeast	96-well plate	>0.1 g/L	L-Tyrosine	91	2 56
109	4-Hydroxystyrene	FUGYGGDSWSUORM-UHFFFAOYSA-N	Yeast	2L tank	>0.1 g/L	L-Tyrosine	91	2 56
110	Resveratrol	LUKBXSAWLPMMSZ-OWOJBTEDSA-N	Yeast	96-well plate	>0.1 g/L	L-Tyrosine	246	3 67
111	Resveratrol	LUKBXSAWLPMMSZ-OWOJBTEDSA-N	Yeast	2L tank	>0.1 g/L	L-Tyrosine	246	3 67
112	1-Hexadecanol	BXWNKGSJHAJOGX-UHFFFAOYSA-N	Yeast	96-well plate	>0.1 g/L	Malonyl-CoA	123	2 55
113	6-Methylsalicylic acid	HCJMNOSIAGSZBM-UHFFFAOYSA-N	Yeast	96-well plate	>0.1 g/L	Malonyl-CoA	155	1 48
114	Phloroglucinol	QCDYQQDYXPDABM-UHFFFAOYSA-N	E. coli	96-well plate	>0.1 g/L	Malonyl-CoA	63	1 37
115	Apigenin	KZNIFHPLKGYRTM-UHFFFAOYSA-N	Yeast	96-well plate	>0.1 g/L	Naringenin	411	5 67
116	Dihydrokaempferol	PADQINQHPQKXNL-LSDHHAIUSA-N	Yeast	96-well plate	>0.1 g/L	Naringenin	392	5 70
117	Genistein	TZBJGXHYKVUXJN-UHFFFAOYSA-N	Yeast	96-well plate	>0.1 g/L	Naringenin	411	6 64
118	Genistein 7-O-beta-D-glucoside	ZCOLJUOHXJRHDI-CMWLGVBASA-N	Yeast	96-well plate	>0.1 g/L	Naringenin	675	7 68
119	Genistein 7-O-beta-D-glucoside	ZCOLJUOHXJRHDI-CMWLGVBASA-N	Yeast	2L tank	>0.1 g/L	Naringenin	675	7 68
120	Kaempferol	IYRMWMYZSQPJKC-UHFFFAOYSA-N	Yeast	96-well plate	>0.1 g/L	Naringenin	451	6 83

No.	Name	InChIKey	Host	Format	Titer	Precursor			
121	Kaempferol	IYRMWMYZSQPJKC-UHFFFAOYSA-N	Yeast	2L tank	>0.1 g/L	Naringenin	451	6	83
122	Quercetin	REFJWTPEDVJJIY-UHFFFAOYSA-N	Yeast	96-well plate	>0.1 g/L	Naringenin	488	7	69
123	Quercetin	REFJWTPEDVJJIY-UHFFFAOYSA-N	Yeast	2L tank	>0.1 g/L	Naringenin	488	7	69
124	Sakuranetin	DJOJDHGQRNZXQQ-AWEZNQCLSA-N	Yeast	96-well plate	>0.1 g/L	Naringenin	377	5	79
125	Sakuranetin	DJOJDHGQRNZXQQ-AWEZNQCLSA-N	Yeast	2L tank	>0.1 g/L	Naringenin	377	5	79
126	Taxifolin	CXQWRCVTCMQVQX-LSDHHAIUSA-N	Yeast	96-well plate	>0.1 g/L	Naringenin	428	6	80
127	Cadaverine	VHRGRCVQAFMJIZ-UHFFFAOYSA-N	Yeast	96-well plate	>0.1 g/L	Oxaloacetate	25	1	63
128	Itaconic acid	LVHBHZANLOWSRM-UHFFFAOYSA-N	Yeast	96-well plate	>0.1 g/L	Oxaloacetate	158	3	60
129	L-Malic Acid	BJEPYKJPYRNKOW-REOHCLBHSA-N	Yeast	96-well plate	>0.1 g/L	Oxaloacetate	129	0	43
130	L-Threonine	AYFVYJQAPQTCCC-GBXIJSLDSA-N	Yeast	96-well plate	>0.1 g/L	Oxaloacetate	93	0	58
131	L-Threonine	AYFVYJQAPQTCCC-GBXIJSLDSA-N	Yeast	2L tank	>0.1 g/L	Oxaloacetate	93	0	58
132	Baicalein	FXNFHKRTJBSTCS-UHFFFAOYSA-N	Yeast	96-well plate	>0.1 g/L	Phenylpyruvate (charge:-1)	413	6	72
133	Baicalein	FXNFHKRTJBSTCS-UHFFFAOYSA-N	Yeast	2L tank	>0.1 g/L	Phenylpyruvate (charge:-1)	413	6	72
134	Phenyllactate	VOXXWSYKYCBWHO-UHFFFAOYSA-N	E. coli	96-well plate	>0.1 g/L	Phenylpyruvate (charge:-1)	150	1	54
135	Styrene oxide	AWMVMTVKBNGEAK-UHFFFAOYSA-N	Yeast	96-well plate	>0.1 g/L	Phenylpyruvate (charge:-1)	95	1	55
136	trans-Cinnamate	WBYWAXJHAXSJNI-VOTSOKGWSA-M	Yeast	96-well plate	>0.1 g/L	Phenylpyruvate (charge:-1)	149	1	65
137	trans-Cinnamate	WBYWAXJHAXSJNI-VOTSOKGWSA-M	Yeast	2L tank	>0.1 g/L	Phenylpyruvate (charge:-1)	149	1	65
138	Lupeol	MQYXUWHLBZFQQO-QGTGJCAVSA-N	Yeast	96-well plate	>0.1 g/L	Squalene	766	2	57
139	(3E)-4,8-dimethylnona-1,3,7-triene	LUKZREJJLWEWQM-YRNVUSSQSA-N	Yeast	96-well plate	>0.1 g/L	trans,trans-Farnesyl diphosphate	167	2	54
140	(E)-gamma-Bisabolene	XBGUIVFBMBVUEG-PFONDFGASA-N	Yeast	2L tank	>0.1 g/L	trans,trans-Farnesyl diphosphate	296	1	56
141	alpha-Copaene	VLXDPFLIRFYIME-UHFFFAOYSA-N	Yeast	96-well plate	>0.1 g/L	trans,trans-Farnesyl diphosphate	312	1	54
142	alpha-Copaene	VLXDPFLIRFYIME-UHFFFAOYSA-N	Yeast	2L tank	>0.1 g/L	trans,trans-Farnesyl diphosphate	312	1	54
143	beta-Selinene	YOVSPTNQHMDJAG-QLFBSQMISA-N	Yeast	96-well plate	>0.1 g/L	trans,trans-Farnesyl diphosphate	286	1	56
144	beta-Sesquiphellandrene	PHWISBHSBNDZDX-LSDHHAIUSA-N	Yeast	96-well plate	>0.1 g/L	trans,trans-Farnesyl diphosphate	264	1	54
145	phlorisobutyrophenone	BNEBXEZRBLYBCZ-UHFFFAOYSA-N	Yeast	96-well plate	>0.01 g/L	(R)-2,3-Dihydroxy-3-methylbutanoate	202	2	56
146	Hydroxyproline	PMMYEEVYMWASQN-UHFFFAOYSA-N	E. coli	96-well plate	>0.01 g/L	2-Oxoglutaric acid	125	3	52
147	L-Glutamine	ZDXPYRJPNDTMRX-VKHMYHEASA-N	Yeast	96-well plate	>0.01 g/L	2-Oxoglutaric acid	146	0	49
148	Spermine	PFNFFQXMRSDOHW-UHFFFAOYSA-N	E. coli	96-well plate	>0.01 g/L	2-Oxoglutaric acid	86	1	96
149	alpha,omega-Dicarboxylic acid	OFOBLEOULBTSOW-UHFFFAOYSA-N	Yeast	96-well plate	>0.01 g/L	Acetyl-CoA	83	1	43
150	alpha,omega-Dicarboxylic acid	OFOBLEOULBTSOW-UHFFFAOYSA-N	Yeast	2L tank	>0.01 g/L	Acetyl-CoA	83	1	43
151	B-methyl-d-valerolactone	YHTLGFCVBKENTE-UHFFFAOYSA-N	Yeast	96-well plate	>0.01 g/L	Acetyl-CoA	99	4	60
152	Prenol	ASUAYTHWZCLXAN-UHFFFAOYSA-N	Yeast	96-well plate	>0.01 g/L	Acetyl-CoA	51	0	48
153	(S)-Norcoclaurine	WZRCQWQRFZITDX-AWEZNQCLSA-N	Yeast	96-well plate	>0.01 g/L	Chorismate	317	4	53
154	4-Hydroxybenzoic acid	FJKROLUGYXJWQN-UHFFFAOYSA-N	Yeast	96-well plate	>0.01 g/L	Chorismate	125	1	54
155	4-Hydroxyphenylacetaldehyde	IPRPPFIAVHPVJH-UHFFFAOYSA-N	Yeast	96-well plate	>0.01 g/L	Chorismate	104	1	52
156	Bacillibactin	RCQTVEFBFUNTGM-BDVHUIKKSA-N	Yeast	96-well plate	>0.01 g/L	Chorismate	1,510	4	74
157	Methyl salicylic acid	OSWPMRLSEDHDFF-UHFFFAOYSA-N	Yeast	96-well plate	>0.01 g/L	Chorismate	144	3	70
158	Methyl salicylic acid	OSWPMRLSEDHDFF-UHFFFAOYSA-N	Yeast	2L tank	>0.01 g/L	Chorismate	144	3	70
159	para-aminobenzoic acid	ALYNCZNDIQEVRV-UHFFFAOYSA-N	Yeast	96-well plate	>0.01 g/L	Chorismate	128	0	46
160	para-aminobenzoic acid	ALYNCZNDIQEVRV-UHFFFAOYSA-N	Yeast	2L tank	>0.01 g/L	Chorismate	128	0	46
161	Phenol	ISWSIDIOOBJBQZ-UHFFFAOYSA-N	Yeast	96-well plate	>0.01 g/L	Chorismate	46	1	50

#	Name	InChIKey	Host	Format	Titer	Precursor			
162	Geraniol	GLZPCOQZEFWAFX-JXMROGBWSA-N	Yeast	96-well plate	>0.01 g/L	geranyl diphosphate (charge:-3)	150	2	55
163	Myrcene	UAHWPYUMFXYFJY-UHFFFAOYSA-N	Yeast	96-well plate	>0.01 g/L	geranyl diphosphate (charge:-3)	145	2	55
164	9beta-Pimara-7,15-diene	VCOVNILQQQZROK-PIKOESSRSA-N	Yeast	96-well plate	>0.01 g/L	Geranylgeranyl diphosphate	441	2	55
165	Canthaxanthin	FDSDTBUPSURDBL-DKLMTRRASA-N	Yeast	96-well plate	>0.01 g/L	Geranylgeranyl diphosphate	1,270	5	52
166	Lycopene	OAIJSZIZWZSQBC-GYZMGTAESA-N	Yeast	96-well plate	>0.01 g/L	Geranylgeranyl diphosphate	1,050	4	57
167	Lycopersene	BGVXBZXEFXMRGJ-DPOFWPLISA-N	Yeast	96-well plate	>0.01 g/L	Geranylgeranyl diphosphate	847	1	67
168	Neoabietadiene	MRRHSEMHYVQUFK-CMKODMSKSA-N	Yeast	96-well plate	>0.01 g/L	Geranylgeranyl diphosphate	458	2	55
169	Retinoic acid	SHGAZHPCJJPHSC-YCNIQYBTSA-N	Yeast	96-well plate	>0.01 g/L	Geranylgeranyl diphosphate	567	5	57
170	Retinoic acid	SHGAZHPCJJPHSC-YCNIQYBTSA-N	Yeast	2L tank	>0.01 g/L	Geranylgeranyl diphosphate	567	5	57
171	Retinol	FPIPGXGPPPQFEQ-OVSJKPMPSA-N	Yeast	96-well plate	>0.01 g/L	Geranylgeranyl diphosphate	496	5	59
172	Retinol	FPIPGXGPPPQFEQ-OVSJKPMPSA-N	Yeast	2L tank	>0.01 g/L	Geranylgeranyl diphosphate	496	5	59
173	Chanoclavine I	SAHHMCVYMGARBT-HEESEWQSSA-O	Yeast	96-well plate	>0.01 g/L	L-Tryptophan	355	3	91
174	Indole-3-acetic acid	SEOVTRFCIGRIMH-UHFFFAOYSA-N	Yeast	96-well plate	>0.01 g/L	L-Tryptophan	205	0	65
175	Indole-3-acetic acid	SEOVTRFCIGRIMH-UHFFFAOYSA-N	Yeast	2L tank	>0.01 g/L	L-Tryptophan	205	0	65
176	pyrrolnitrin	QJBZDBLBQWFTPZ-UHFFFAOYSA-N	E. coli	96-well plate	>0.01 g/L	L-Tryptophan	272	4	68
177	pyrrolnitrin	QJBZDBLBQWFTPZ-UHFFFAOYSA-N	E. coli	96-well plate	>0.01 g/L	L-Tryptophan	272	4	68
178	Tryptamine	APJYDQYYACXCRM-UHFFFAOYSA-N	Yeast	96-well plate	>0.01 g/L	L-Tryptophan	147	1	69
179	Tryptamine	APJYDQYYACXCRM-UHFFFAOYSA-N	Yeast	2L tank	>0.01 g/L	L-Tryptophan	147	1	69
180	Bisdemethoxycurcumin	PREBVFJICNPEKM-YDWXAUTNSA-N	Yeast	96-well plate	>0.01 g/L	L-Tyrosine	408	3	68
181	Bisdemethoxycurcumin	PREBVFJICNPEKM-YDWXAUTNSA-N	Yeast	2L tank	>0.01 g/L	L-Tyrosine	408	3	68
182	p-Hydroxybenzyl acetone	NJGBTKGETPDVIK-UHFFFAOYSA-N	Yeast	96-well plate	>0.01 g/L	L-Tyrosine	146	4	67
183	Fatty Acid Ethyl Esters	XIRNKXNNONJFQO-UHFFFAOYSA-N	Yeast	96-well plate	>0.01 g/L	Malonyl-CoA	202	1	63
184	hexanol	ZSIAUFGUXNUGDI-UHFFFAOYSA-N	Yeast	96-well plate	>0.01 g/L	Malonyl-CoA	27	4	59
185	(+)-Catechin	PFTAWBLQPZVEMU-DZGCQCFKSA-N	Yeast	96-well plate	>0.01 g/L	Naringenin	364	8	80
186	7-O-Methyl Aromadendrin	LZLGHWHSUZVUFZ-UHFFFAOYSA-N	Yeast	96-well plate	>0.01 g/L	Naringenin	405	6	88
187	Apiforol	RPKUCYSGAXIESU-ABLWVSNPSA-N	Yeast	96-well plate	>0.01 g/L	Naringenin	328	5	61
188	Apiforol	RPKUCYSGAXIESU-ABLWVSNPSA-N	Yeast	96-well plate	>0.01 g/L	Naringenin	328	5	61
189	Cyanidin	VEVZSMAEJFVWIL-UHFFFAOYSA-O	Yeast	96-well plate	>0.01 g/L	Naringenin	364	8	84
190	Cyanidin	VEVZSMAEJFVWIL-UHFFFAOYSA-O	Yeast	2L tank	>0.01 g/L	Naringenin	364	8	84
191	Chrysin	RTIXKCRFFJGDFG-UHFFFAOYSA-N	Yeast	96-well plate	>0.01 g/L	Phenylpyruvate (charge:-1)	384	5	71
192	Chrysin	RTIXKCRFFJGDFG-UHFFFAOYSA-N	Yeast	2L tank	>0.01 g/L	Phenylpyruvate (charge:-1)	384	5	71
193	L-Phenylalanine	COLNVLDHVKWLRT-QMMMGPOBSA-N	Yeast	96-well plate	>0.01 g/L	Phenylpyruvate (charge:-1)	153	0	60
194	L-Phenylalanine	COLNVLDHVKWLRT-QMMMGPOBSA-N	Yeast	2L tank	>0.01 g/L	Phenylpyruvate (charge:-1)	153	0	60
195	Pinosylvin	YCVPRTHEGLPYPB-VOTSOKGWSA-N	Yeast	96-well plate	>0.01 g/L	Phenylpyruvate (charge:-1)	221	3	64
196	Pinosylvin	YCVPRTHEGLPYPB-VOTSOKGWSA-N	Yeast	2L tank	>0.01 g/L	Phenylpyruvate (charge:-1)	221	3	64
197	Glycyrrhetinic acid	MPDGHEJMBKOTSU-YKLVYJNSSA-N	Yeast	96-well plate	>0.01 g/L	Squalene	965	4	63
198	Protopanaxadiol	PYXFVCFISTUSOO-VUFVRDRTSA-N	Yeast	96-well plate	>0.01 g/L	Squalene	783	3	55
199	(-)-beta-elemene	OPFTUNCRGUEPRZ-QLFBSQMISA-N	Yeast	96-well plate	>0.01 g/L	trans,trans-Farnesyl diphosphate	284	1	54
200	(+)-epi-Isozizaene	CYLSPJUZBPWJGC-ITDIGPHOSA-N	Yeast	96-well plate	>0.01 g/L	trans,trans-Farnesyl diphosphate	334	1	54
201	(2E,6E)-Farnesol	CRDAMVZIKSXKFV-YFVJMOTDSA-N	Yeast	96-well plate	>0.01 g/L	trans,trans-Farnesyl diphosphate	265	1	44
202	(E)-gamma-Bisabolene	XBGUIVFBMBVUEG-PFONDFGASA-N	Yeast	96-well plate	>0.01 g/L	trans,trans-Farnesyl diphosphate	296	1	56
203	5-Epiaristolochene	YONHOSLUBQJXPR-UMVBOHGHSA-N	Yeast	96-well plate	>0.01 g/L	trans,trans-Farnesyl diphosphate	297	1	54

204	7-epi-Sesquithujene	UCQHFDKBUHCAFR-KKUMJFAQSA-N	Yeast	96-well plate	>0.01 g/L	trans,trans-Farnesyl diphosphate	304	1	54
205	beta-Cubebene	FSRZGYRCMPZNJF-KHMAMNHCSA-N	Yeast	96-well plate	>0.01 g/L	trans,trans-Farnesyl diphosphate	301	1	54
206	beta-Cubebene	FSRZGYRCMPZNJF-KHMAMNHCSA-N	Yeast	2L tank	>0.01 g/L	trans,trans-Farnesyl diphosphate	301	1	54
207	beta-Eudesmol	BOPIMTNSYWYZOC-VNHYZAJKSA-N	Yeast	96-well plate	>0.01 g/L	trans,trans-Farnesyl diphosphate	292	1	54
208	beta-Eudesmol	BOPIMTNSYWYZOC-VNHYZAJKSA-N	Yeast	2L tank	>0.01 g/L	trans,trans-Farnesyl diphosphate	292	1	54
209	Elemol	GFJIQNADMLPFOW-VNHYZAJKSA-N	Yeast	96-well plate	>0.01 g/L	trans,trans-Farnesyl diphosphate	290	1	58
210	Longifolene	PDSNLYSELAIEBU-UHFFFAOYSA-N	Yeast	96-well plate	>0.01 g/L	trans,trans-Farnesyl diphosphate	312	1	54
211	Longifolene	PDSNLYSELAIEBU-UHFFFAOYSA-N	Yeast	2L tank	>0.01 g/L	trans,trans-Farnesyl diphosphate	312	1	54
212	Valerena-4,7(11)-diene	MZZFDMZYIBWOOA-KWCYVHTRSA-N	Yeast	96-well plate	>0.01 g/L	trans,trans-Farnesyl diphosphate	302	1	62
213	Valerena-4,7(11)-diene	MZZFDMZYIBWOOA-KWCYVHTRSA-N	Yeast	2L tank	>0.01 g/L	trans,trans-Farnesyl diphosphate	302	1	62
214	Vetispiradiene	WEZDOYDDKIHCLM-QLFBSQMISA-N	Yeast	96-well plate	>0.01 g/L	trans,trans-Farnesyl diphosphate	297	1	54
215	Hyaluronic acid	KIUKXJAPPMFGSW-MNSSHETKSA-N	Yeast	96-well plate	>0.001 g/L	2-Oxoglutaric acid	1,300	3	54
216	caffeic acid	QAIPRVGONGVQAS-DUXPYHPUSA-N	Yeast	96-well plate	>0.001 g/L	3,4-Dihydroxy-L-phenylalanine	212	2	50
217	o-Methoxyphenol	LHGVFZTZFXWLCP-UHFFFAOYSA-N	Yeast	96-well plate	>0.001 g/L	3-Dehydroshikimate	83	3	66
218	1,3-Dihydroxy-N-methylacridone	GDALETGZDYOOGB-UHFFFAOYSA-N	Yeast	96-well plate	>0.001 g/L	Chorismate	346	3	80
219	(6E)-8-Hydroxygeraniol	PREUOUJFXMCMSJ-TXFIJWAUSA-N	Yeast	96-well plate	>0.001 g/L	geranyl diphosphate (charge:-3)	169	3	51
220	(+)-Sandaracopimaradiene	XDSYKASBVOZOAG-QGZVKYPTSA-N	Yeast	96-well plate	>0.001 g/L	Geranylgeranyl diphosphate	441	2	55
221	(E,E)-Geranyllinalool	IQDXAJNQKSIPGB-HQSZAHFGSA-N	Yeast	96-well plate	>0.001 g/L	Geranylgeranyl diphosphate	394	1	57
222	beta-Ionone	PSQYTAPXSHCGMF-BQYQJAHWSA-N	Yeast	96-well plate	>0.001 g/L	Geranylgeranyl diphosphate	292	6	55
223	ent-Cassa-12,15-diene	JQPDOKGAOXSRJD-SVEODPQUSA-N	Yeast	96-well plate	>0.001 g/L	Geranylgeranyl diphosphate	427	2	55
224	Lutein	KBPHJBAIARWVSC-RGZFRNHPSA-N	Yeast	96-well plate	>0.001 g/L	Geranylgeranyl diphosphate	1,270	6	57
225	Choline	SGMZJAMFUVOLNK-UHFFFAOYSA-M	Yeast	96-well plate	>0.001 g/L	L-Serine	47	0	74
226	Glycine	DHMQDGOQFOQNFH-UHFFFAOYSA-N	Yeast	96-well plate	>0.001 g/L	L-Serine	43	0	57
227	L-Methionine	FFEARJCKVFRZRR-BYPYZUCNSA-N	Yeast	96-well plate	>0.001 g/L	L-Serine	97	0	78
228	Phytosphingosine	AERBNCYCJBRYDG-KSZLIROESA-N	Yeast	96-well plate	>0.001 g/L	L-Serine	226	0	67
229	4-(3-Methylbut-2-enyl)-L-tryptophan	MZSPRSJAOSKAAT-ZDUSSCGKSA-N	Yeast	96-well plate	>0.001 g/L	L-Tryptophan	371	1	79
230	Curcumin	VFLDPWHFBUODDF-FCXRPNKRSA-N	Yeast	96-well plate	>0.001 g/L	L-Tyrosine	507	6	79
231	Ferulic acid	KSEBMYQBYZTDHS-HWKANZROSA-N	Yeast	96-well plate	>0.001 g/L	L-Tyrosine	224	3	76
232	Liquiritigenin	FURUXTVZLHCCNA-AWEZNQCLSA-N	Yeast	96-well plate	>0.001 g/L	L-Tyrosine	335	4	65
233	Synephrine	YRCWQPVGYLYSOX-UHFFFAOYSA-N	Yeast	96-well plate	>0.001 g/L	L-Tyrosine	122	4	80
234	Dihydromonacolin L	AGNDLYBQPUJADV-VCWNUMGPSA-N	Yeast	96-well plate	>0.001 g/L	Malonyl-CoA	430	1	69
235	Apigenin 7-O-beta-D-glucoside	KMOUJOKENFFTPU-QNDFHXLGSA-N	Yeast	96-well plate	>0.001 g/L	Naringenin	675	6	68
236	Delphinidin	JKHRCGUTYDNCLE-UHFFFAOYSA-O	Yeast	96-well plate	>0.001 g/L	Naringenin	380	8	86
237	Eriodictyol	SBHXYTNGIZCORC-ZDUSSCGKSA-N	Yeast	96-well plate	>0.001 g/L	Naringenin	400	5	64
238	Eriodictyol	SBHXYTNGIZCORC-ZDUSSCGKSA-N	Yeast	2L tank	>0.001 g/L	Naringenin	400	5	64
239	Genkwanin	JPMYFOBNRRGFNO-UHFFFAOYSA-N	Yeast	96-well plate	>0.001 g/L	Naringenin	424	6	82
240	Hesperidin	QUUPHWDTPGMPEX-QJBIFVCTSA-N	Yeast	96-well plate	>0.001 g/L	Naringenin	940	9	85
241	Myricetin	IKMDFBPHZNJCSN-UHFFFAOYSA-N	Yeast	96-well plate	>0.001 g/L	Naringenin	506	7	83
242	Pelargonidin	XVFMGWDSJLBXDZ-UHFFFAOYSA-O	Yeast	96-well plate	>0.001 g/L	Naringenin	328	7	76
243	Pelargonidin	XVFMGWDSJLBXDZ-UHFFFAOYSA-O	Yeast	2L tank	>0.001 g/L	Naringenin	328	7	76
244	L-Lysine	KDXKERNSBIXSRK-YFKPBYRVSA-N	Yeast	96-well plate	>0.001 g/L	Oxaloacetate	106	0	66
245	Pinocembrin	URFCJEUYXNAHFI-ZDUSSCGKSA-N	Yeast	96-well plate	>0.001 g/L	Phenylpyruvate (charge:-1)	337	4	66

246	Pinocembrin	URFCJEUYXNAHFI-ZDUSSCGKSA-N	Yeast	2L tank	>0.001 g/L	Phenylpyruvate (charge:-1)	337	4 66
247	Soyasapogenol B	YOQAQNKGFOLRGT-UXXABWCISA-N	Yeast	96-well plate	>0.001 g/L	Squalene	848	6 59
248	(+)-delta-Selinene	VEGYMPQCXPVQJY-OAHLLOKOSA-N	Yeast	96-well plate	>0.001 g/L	trans,trans-Farnesyl diphosphate	317	1 54
249	Cedrene	IRAQOCYXUMOFCW-OSFYFWSMSA-N	Yeast	96-well plate	>0.001 g/L	trans,trans-Farnesyl diphosphate	323	1 54
250	Cubebol	KONGRWVLXLWGDV-BYGOPZEFSA-N	Yeast	96-well plate	>0.001 g/L	trans,trans-Farnesyl diphosphate	311	1 54
251	Zingiberene	KKOXKGNSUHTUBV-LSDHHAIUSA-N	Yeast	96-well plate	>0.001 g/L	trans,trans-Farnesyl diphosphate	274	1 54
252	Carnosine	CQOVPNPJLQNMDC-ZETCQYMHSA-N	Yeast	96-well plate	Detected (unquantified)	(R)-2,3-Dihydroxy-3-methylbutanoate	259	1 96
253	Columbamine	YYFOFDHQVIODOQ-UHFFFAOYSA-O	Yeast	96-well plate	Detected (unquantified)	(S)-reticuline (charge:1)	461	11 80
254	Salutaridinol	LLSADFZHWMEBHH-TYILLQQXSA-N	Yeast	96-well plate	Detected (unquantified)	(S)-reticuline (charge:1)	574	11 73
255	Stylopine	UXYJCYXWJGAKQY-UHFFFAOYSA-N	Yeast	96-well plate	Detected (unquantified)	(S)-reticuline (charge:1)	502	11 76
256	D-Sorbitol	FBPFZTCFMRRESA-JGWLITMVSA-N	Yeast	96-well plate	Detected (unquantified)	[NULL]	105	2 17
257	4-Aminobutanoic acid	BTCSSZJGUNDROE-UHFFFAOYSA-N	Yeast	96-well plate	Detected (unquantified)	2-Oxoglutaric acid	63	0 49
258	N-Acetyl-D-glucosamine	MBLBDJOUHNCFQT-LXGUWJNJSA-N	Yeast	96-well plate	Detected (unquantified)	2-Oxoglutaric acid	221	1 51
259	(9Z)-Octadecenoic acid	ZQPPMHVWECSIRJ-KTKRTIGZSA-N	Yeast	96-well plate	Detected (unquantified)	Acetyl-CoA	234	0 68
260	epsilon-Caprolactam	JBKVHLHDHHXQEQ-UHFFFAOYSA-N	Yeast	96-well plate	Detected (unquantified)	Acetyl-CoA	91	7 61
261	Zeatin	UZKQTCBAMSWPJD-FARCUNLSSA-N	Yeast	96-well plate	Detected (unquantified)	Acetyl-CoA	258	4 82
262	L-Histidine	HNDVDQJCIGZPNO-YFKPBYRVSA-N	Yeast	96-well plate	Detected (unquantified)	alpha-D-Ribose 1-phosphate	151	0 84
263	Pyridoxine	LXNHXLLTXMVWPM-UHFFFAOYSA-N	Yeast	96-well plate	Detected (unquantified)	alpha-D-Ribose 1-phosphate	142	2 51
264	Theophylline	ZFXYFBGIUFBOJW-UHFFFAOYSA-N	Yeast	96-well plate	Detected (unquantified)	alpha-D-Ribose 1-phosphate	267	2 89
265	(S)-Coclaurine	LVVKXRQZSRUVPY-HNNXBMFYSA-O	Yeast	96-well plate	Detected (unquantified)	Chorismate	330	5 74
266	2,4,6-Trihydroxybenzophenone	CPEXFJVZFNYXGU-UHFFFAOYSA-N	Yeast	96-well plate	Detected (unquantified)	Chorismate	261	5 63
267	Acetoacetic acid	WDJHALXBUFZDSR-UHFFFAOYSA-N	Yeast	96-well plate	Detected (unquantified)	Chorismate	95	0 58
268	Fumiquinazoline	POEYRUBMWIOMTB-NYGOOMQSSA-N	Yeast	96-well plate	Detected (unquantified)	Chorismate	984	4 71
269	Isochorismate	NTGWPRCCOQCMGE-YUMQZZPRSA-N	Yeast	96-well plate	Detected (unquantified)	Chorismate	392	1 55
270	(-)-alpha-Terpineol	WUOACPNHFRMFPN-SECBINFHSA-N	Yeast	96-well plate	Detected (unquantified)	geranyl diphosphate (charge:-3)	168	2 51
271	(-)-endo-Fenchol	IAIHUHQCLTYTSF-MRTMQBJTSA-N	Yeast	96-well plate	Detected (unquantified)	geranyl diphosphate (charge:-3)	185	1 61
272	(+)-Camphor	DSSYKIVIOFKYAU-UHFFFAOYSA-N	Yeast	96-well plate	Detected (unquantified)	geranyl diphosphate (charge:-3)	217	4 51
273	(+)-Linalool	CDOSHBSSFJOMGT-SNVBAGLBSA-N	Yeast	96-well plate	Detected (unquantified)	geranyl diphosphate (charge:-3)	154	2 51
274	Geranial	WTEVQBCEXWBHNA-JXMROGBWSA-N	Yeast	96-well plate	Detected (unquantified)	geranyl diphosphate (charge:-3)	171	3 51
275	Loganate	JNNGEAWILNVFFD-CDJYTOATSA-M	Yeast	96-well plate	Detected (unquantified)	geranyl diphosphate (charge:-3)	560	8 53

33

276	beta-Springene	XSIVJVJUIXOEPW-YTOAGJIJSA-N	Yeast	96-well plate	Detected (unquantified)	Geranylgeranyl diphosphate	384	2	60
277	Casbene	ZJMVJDFTNPZVMB-QOCMWZQCSA-N	Yeast	96-well plate	Detected (unquantified)	Geranylgeranyl diphosphate	431	1	55
278	Isopimaric acid	MXYATHGRPJZBNA-KRFUXDQASA-N	Yeast	96-well plate	Detected (unquantified)	Geranylgeranyl diphosphate	534	6	66
279	Retinal	NCYCYZXNIZJOKI-OVSJKPMPSA-N	Yeast	96-well plate	Detected (unquantified)	Geranylgeranyl diphosphate	522	4	58
280	5-Aminolevulinic acid	ZGXJTSGNIOSYLO-UHFFFAOYSA-N	E. coli	96-well plate	Detected (unquantified)	L-Serine	121	0	48
281	S-Methyl-L-methionine	YDBYJHTYSHBBAU-YFKPBYRVSA-O	Yeast	96-well plate	Detected (unquantified)	L-Serine	116	1	85
282	violacein	XAPNKXIRQFHCHN-QGOAFFKASA-N	Yeast	96-well plate	Detected (unquantified)	L-Tryptophan	713	7	69
283	Daidzein	ZQSIJRDFPHDXIC-UHFFFAOYSA-N	Yeast	96-well plate	Detected (unquantified)	L-Tyrosine	382	5	70
284	Daidzin	KYQZWONCHDNPDP-QNDFHXLGSA-N	Yeast	96-well plate	Detected (unquantified)	L-Tyrosine	644	6	70
285	Isoliquiritigenin	DXDRHHKMWQZJHT-FPYGCLRLSA-N	E. coli	96-well plate	Detected (unquantified)	L-Tyrosine	331	3	60
286	Phenethylamine	BHHGXPLMPWCGHP-UHFFFAOYSA-N	Yeast	96-well plate	Detected (unquantified)	L-Tyrosine	65	1	49
287	Thiamine	JZRWCGZRTZMZEH-UHFFFAOYSA-N	Yeast	96-well plate	Detected (unquantified)	L-Tyrosine	269	0	105
288	2,4-Dihydroxy-6-pentylbenzoic acid	SXFKFRRXJUJGSS-UHFFFAOYSA-N	Yeast	96-well plate	Detected (unquantified)	Malonyl-CoA	229	5	53
289	3-Hydroxypropionic Acid	ALRHLSYJTWAHJZ-UHFFFAOYSA-N	Yeast	96-well plate	Detected (unquantified)	Malonyl-CoA	50	2	55
290	Lovastatin	PCZOHLXUXFIOCF-BXMDZJJMSA-N	Yeast	96-well plate	Detected (unquantified)	Malonyl-CoA	666	7	69
291	Monacolin J	ZDFOBOYQVYMVCW-IRUSZSJRSA-N	Yeast	96-well plate	Detected (unquantified)	Malonyl-CoA	510	5	73
292	cyanidin 3-O-beta-D-glucoside (charge:-1)	RKWHWFONKJEUEF-GQUPQBGVSA-M	Yeast	96-well plate	Detected (unquantified)	Naringenin	623	9	77
293	Luteolin	IQPNAANSBPBGFQ-UHFFFAOYSA-N	Yeast	96-well plate	Detected (unquantified)	Naringenin	447	6	67
294	Ponciretin	HMUJXQRRKBLVOO-UHFFFAOYSA-M	Yeast	96-well plate	Detected (unquantified)	Naringenin	377	5	79
295	(R,R)-Tartaric acid	FEWJPZIEWOKRBE-JCYAYHJZSA-N	Yeast	96-well plate	Detected (unquantified)	Oxaloacetate	134	1	41
296	(S,S)-Tartaric acid	FEWJPZIEWOKRBE-LWMBPPNESA-N	Yeast	96-well plate	Detected (unquantified)	Oxaloacetate	134	1	41
297	Glyoxylate	HHLFWLYXYJOTON-UHFFFAOYSA-M	Yeast	96-well plate	Detected (unquantified)	Oxaloacetate	50	0	44
298	Hydroxymalonic acid	ROBFUDYVXSDBQM-UHFFFAOYSA-N	Yeast	96-well plate	Detected (unquantified)	Oxaloacetate	103	4	49
299	1,2-phenylethanediol	PWMWNFMRSKOCEY-UHFFFAOYSA-N	Yeast	96-well plate	Detected (unquantified)	Phenylpyruvate (charge:-1)	87	3	62
300	1,2-phenylethanediol	PWMWNFMRSKOCEY-UHFFFAOYSA-N	Yeast	2L tank	Detected (unquantified)	Phenylpyruvate (charge:-1)	87	3	62
301	Anethole	RUVINXPYWBROJD-UHFFFAOYSA-N	Yeast	96-well plate	Detected (unquantified)	Phenylpyruvate (charge:-1)	121	8	68
302	Isopenicillin N	MIFYHUACUWQUKT-GTQWGBSQSA-N	Yeast	96-well plate	Detected (unquantified)	Phenylpyruvate (charge:-1)	581	2	82
303	Phenylpyruvic Acid	BTNMPGBKDVTSJY-UHFFFAOYSA-N	Yeast	96-well plate	Detected (unquantified)	Phenylpyruvate (charge:-1)	180	0	50

	Name	InChIKey	Organism	Format	Detection	Precursor	Value	N	N2
304	Acetoin	ROWKJAVDOGWPAT-UHFFFAOYSA-N	Yeast	96-well plate	Detected (unquantified)	Pyruvate	59	0	23
305	Glutaric Acid	JFCQEDHGNNZCLN-UHFFFAOYSA-N	Yeast	96-well plate	Detected (unquantified)	Pyruvate	104	3	49
306	L-Isoleucine	AGPKZVBTJJNPAG-WHFBIAKZSA-N	Yeast	96-well plate	Detected (unquantified)	Pyruvate	103	0	71
307	Propionic Acid	XBDQKXXYIPTUBI-UHFFFAOYSA-N	Yeast	96-well plate	Detected (unquantified)	Pyruvate	40	2	57
308	Alpha-Amyrin	FSLPMRQHCOLESF-SFMCKYFRSA-N	Yeast	96-well plate	Detected (unquantified)	Squalene	779	2	59
309	beta-Amyrin	JFSHUTJDVKUMTJ-QHPUVITPSA-N	Yeast	96-well plate	Detected (unquantified)	Squalene	790	2	55
310	Dammarenediol II	NLHQJXWYMZLQJY-TXNIMPHESA-N	Yeast	96-well plate	Detected (unquantified)	Squalene	750	2	55
311	Lanosterol	CAHGCLMLTWQZNJ-BQNIITSRSA-N	Yeast	96-well plate	Detected (unquantified)	Squalene	767	0	73
312	Oleanolic acid	MIJYXULNPSFWEK-GTOFXWBISA-N	Yeast	96-well plate	Detected (unquantified)	Squalene	885	3	61
313	Thalianol	DGAGPZOBTQYNRE-VMSIWEJCSA-N	Yeast	96-well plate	Detected (unquantified)	Squalene	746	2	59
314	8-epi-Cedrol	SVURIXNDRWRAFU-SUWDNJBVSA-N	Yeast	96-well plate	Detected (unquantified)	trans,trans-Farnesyl diphosphate	321	1	56
315	Albaflavenone	SHUZZAXJEJPUGA-CCUNJIBTSA-N	Yeast	96-well plate	Detected (unquantified)	trans,trans-Farnesyl diphosphate	402	2	54
316	Costunolide	HRYLQFBHBWLLLL-AHNJNIBGSA-N	Yeast	96-well plate	Detected (unquantified)	trans,trans-Farnesyl diphosphate	401	4	58
317	Nootkatone	WTOYNNBCKUYIKC-JMSVASOKSA-N	Yeast	96-well plate	Detected (unquantified)	trans,trans-Farnesyl diphosphate	364	3	62
318	2-Aminoethylphosphonic Acid	QQVDJLLNRSOCEL-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	(R)-2,3-Dihydroxy-3-methylbutanoate	87	2	44
319	3-Methyl-1-Butanol	PHTQWCKDNZKARW-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	(R)-2,3-Dihydroxy-3-methylbutanoate	25	0	49
320	4-Methyl-2-oxopentanoic acid	BKAJNAXTPSGJCU-UHFFFAOYSA-N	E. coli	96-well plate	Not detected above host organism background levels	(R)-2,3-Dihydroxy-3-methylbutanoate	126	0	45
321	Butyric Acid	FERIUCNNQQJTOY-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	(R)-2,3-Dihydroxy-3-methylbutanoate	50	3	46
322	Isobutanol	ZXEKIIBDNHEJCQ-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	(R)-2,3-Dihydroxy-3-methylbutanoate	18	0	44
323	L-Leucine	ROHFNLRQFUQHCH-YFKPBYRVSA-N	Yeast	96-well plate	Not detected above host organism background levels	(R)-2,3-Dihydroxy-3-methylbutanoate	101	0	61
324	(R)-Reticuline	BHLYRWXGMIUIHG-OAHLLOKOSA-N	Yeast	96-well plate	Not detected above host organism background levels	(S)-reticuline (charge:1)	407	9	80
325	(S)-Canadine	VZTUIEROBZXUFA-INIZCTEOSA-N	Yeast	96-well plate	Not detected above host organism background levels	(S)-reticuline (charge:1)	488	11	77
326	(S)-Reticuline	BHLYRWXGMIUIHG-HNNXBMFYSA-N	Yeast	96-well plate	Not detected above host organism	(S)-reticuline (charge:1)	407	8	79

No.	Name	InChIKey	Organism	Plate	Detection	Precursor			
					background levels				
327	(S)-Scoulerine	KNWVMRVOBAFFMH-HNNXBMFYSA-N	Yeast	96-well plate	Not detected above host organism background levels	(S)-reticuline (charge:1)	447	9	77
328	Berberine	YBHILYKTIRIUTE-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	(S)-reticuline (charge:1)	488	12	76
329	Cheilanthifoline	FVXCQULKSPVRPK-HNNXBMFYSA-N	Yeast	96-well plate	Not detected above host organism background levels	(S)-reticuline (charge:1)	474	10	83
330	Isocorypalmine	KDFKJOFJHSVROC-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	(S)-reticuline (charge:1)	461	10	76
331	Palmatine	QUCQEUCGKKTEBI-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	(S)-reticuline (charge:1)	475	12	75
332	Protopine	GPTFURBXHJWNHR-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	(S)-reticuline (charge:1)	542	13	83
333	Salutaridine	GVTRUVGBZQJVTF-YJYMSZOUSA-N	Yeast	96-well plate	Not detected above host organism background levels	(S)-reticuline (charge:1)	613	10	73
334	Sanguinarine	INVGWHRKADIJHF-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	(S)-reticuline (charge:1)	530	16	75
335	Ascorbic acid	CIWBSHSKHKDKBQ-JLAZNSOCSA-N	Yeast	96-well plate	Not detected above host organism background levels	[NULL]	232	4	47
336	Erythritol	UNXHWFMMPAWVPI-ZXZARUISSA-N	Yeast	96-well plate	Not detected above host organism background levels	[NULL]	48	2	52
337	Gluconic Acid	RGHNJXZEOKUKBD-SQOUGZDYSA-N	Yeast	96-well plate	Not detected above host organism background levels	[NULL]	170	1	45
338	Kojic Acid	BEJNERDRQOWKJM-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	[NULL]	214	3	37
339	1,3-Diaminopropane	XFNJVJPLKCPIBV-UHFFFAOYSA-N	E. coli	96-well plate	Not detected above host organism background levels	2-Oxoglutaric acid	12	1	98
340	1,4 Butanediol (Bdo)	WERYXYBDKMZEQL-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	2-Oxoglutaric acid	18	4	58
341	5-Aminopentanoic acid	JJMDCOVWQOJGCB-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	2-Oxoglutaric acid	73	2	61
342	6-Aminohexanoic Acid	SLXKOJJOQWFEFD-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	2-Oxoglutaric acid	83	6	61
343	Clavaminic acid	GQHALSXZONOXGJ-WHJCQOFKSA-N	E. coli	96-well plate	Not detected above host organism background levels	2-Oxoglutaric acid	325	6	63

344	Deacetoxycephalosporin C	NNQIJOYQWYKBOW-JWKOBGCHSA-N	Yeast	96-well plate	Not detected above host organism background levels	2-Oxoglutaric acid	620	4	86
345	ferrichrome	GGUNGDGGXMHBMJ-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	2-Oxoglutaric acid	1,200	4	76
346	Fr-33289	SMXUDUOWTDBXDT-RXMQYKEDSA-M	Yeast	96-well plate	Not detected above host organism background levels	2-Oxoglutaric acid	223	12	64
347	Fr-900098	PKMNDDZSIHLLLI-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	2-Oxoglutaric acid	200	11	59
348	L-Arginine	ODKSFYDXXFIFQN-BYPYZUCNSA-N	Yeast	96-well plate	Not detected above host organism background levels	2-Oxoglutaric acid	176	0	70
349	L-Proline	ONIBWKKTOPOVIA-BYPYZUCNSA-N	Yeast	96-well plate	Not detected above host organism background levels	2-Oxoglutaric acid	103	0	62
350	Plumbemycin A	BMFVTRZNBFRUMH-NSCUHMNNSA-N	Yeast	96-well plate	Not detected above host organism background levels	2-Oxoglutaric acid	601	7	55
351	Putrescine	KIDHWZJUCRJVML-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	2-Oxoglutaric acid	18	0	54
352	sym-Homospermidine	UODZHRGDSPLRMD-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	2-Oxoglutaric acid	58	1	59
353	Tropinone	QQXLDOJGLXJCSE-UHFFFAOYSA-N	E. coli	96-well plate	Not detected above host organism background levels	2-Oxoglutaric acid	151	7	126
354	(R)-rosmarinic acid	DOUMFZQKYFQNTF-WUTVXBCWSA-N	Yeast	96-well plate	Not detected above host organism background levels	3,4-Dihydroxy-L-phenylalanine	519	6	58
355	3-(3,4-Dihydroxyphenyl)Lactate	PAFLSMZLRSPALU-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	3,4-Dihydroxy-L-phenylalanine	205	3	60
356	3-(3,4-Dihydroxyphenyl)pyruvate	LQQFFJFGLSKYIR-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	3,4-Dihydroxy-L-phenylalanine	237	2	58
357	Betanin	DHHFDKNIEVKVKS-FMOSSLLZSA-N	Yeast	96-well plate	Not detected above host organism background levels	3,4-Dihydroxy-L-phenylalanine	1,070	7	70
358	Epinephrine	UCTWMZQNUQWSLP-VIFPVBQESA-N	Yeast	96-well plate	Not detected above host organism background levels	3,4-Dihydroxy-L-phenylalanine	154	5	73
359	Eugenol	RRAFCDWBNXTKKO-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	3,4-Dihydroxy-L-phenylalanine	145	8	75
360	Indole-5,6-quinone	IGGVVGHJSQSLFO-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	3,4-Dihydroxy-L-phenylalanine	334	6	59
361	L-Dopachrome	VJNCICVKUHKIIV-LURJTMIESA-N	Yeast	96-well plate	Not detected above host	3,4-Dihydroxy-L-phenylalanine	405	4	57

					organism background levels			
362	L-Noradrenaline	SFLSHLFXELFNJZ-QMMMGPOBSA-N	Yeast	96-well plate	Not detected above host organism background levels	3,4-Dihydroxy-L-phenylalanine	142	4 58
363	Capsaicin	YKPUWZUDDOIDPM-SOFGYWHQSA-N	Yeast	96-well plate	Not detected above host organism background levels	3-Dehydroshikimate	341	7 125
364	cis,cis-Muconate	TXXHDPDFNKHHGW-CCAGOZQPSA-N	Yeast	96-well plate	Not detected above host organism background levels	3-Dehydroshikimate	168	2 53
365	Gallic acid	LNTHITQWFMADLM-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	3-Dehydroshikimate	169	2 51
366	1-heptanol	BBMCTIGTTCKYKF-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	Acetyl-CoA	35	7 52
367	1-Octanol	KBPLFHHGFOOTCA-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	Acetyl-CoA	44	2 54
368	1-Octen-3-Ol	VSMOENVRRABVKN-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	Acetyl-CoA	69	7 62
369	1-Pentanol	AMQJEAYHLZJPGS-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	Acetyl-CoA	20	4 63
370	2-Methyl-1-Butanol	QPRQEDXDYOZYLA-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	Acetyl-CoA	27	0 65
371	2-Methyl-3-Buten-2-Ol	HNVRRHSXBLFLIG-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	Acetyl-CoA	55	1 52
372	2-methylpropanedioate	ZIYVHBGGAOATLY-UHFFFAOYSA-L	Yeast	96-well plate	Not detected above host organism background levels	Acetyl-CoA	92	2 82
373	6'S,10'R)-7',8'-Dehydrozearalenol	NAXNFNYRXNVLQZ-NZUVUHHCSA-N	Yeast	96-well plate	Not detected above host organism background levels	Acetyl-CoA	445	1 49
374	Adipate Semialdehyde	PNPPVRALIYXJBW-UHFFFAOYSA-M	Yeast	96-well plate	Not detected above host organism background levels	Acetyl-CoA	93	5 61
375	Adipic acid	WNLRTRBMVRJNCN-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	Acetyl-CoA	114	4 58
376	Aigialomycin C	IDIVCUJKNPAADU-JUEMZMRLSA-N	Yeast	96-well plate	Not detected above host organism background levels	Acetyl-CoA	522	2 66
377	Arachidonic Acid	YZXBAPSDXZZRGB-DOFZRALJSA-N	Yeast	96-well plate	Not detected above host organism background levels	Acetyl-CoA	362	1 55
378	Eicosapentaenoic Acid	JAZBEHYOTPTENJ-JLNKQSITSA-N	Yeast	96-well plate	Not detected above host organism	Acetyl-CoA	398	1 56

38

					background levels			
379	Gamma-Linolenic Acid	VZCCETWTMQHEPK-QNEBEIHSSA-N	Yeast	96-well plate	Not detected above host organism background levels	Acetyl-CoA	301	6 68
380	Hypericin	BTXNYTINYBABQR-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	Acetyl-CoA	981	4 49
381	Isoprenol	CPJRRXSHAYUTGL-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	Acetyl-CoA	48	2 61
382	Linoleic acid	OYHQOLUKZRVURQ-HZJYTTRNSA-N	Yeast	96-well plate	Not detected above host organism background levels	Acetyl-CoA	267	5 67
383	Linolenic Acid	DTOSIQBPPRVQHS-PDBXOOCHSA-N	Yeast	96-well plate	Not detected above host organism background levels	Acetyl-CoA	301	6 67
384	Prostacyclin	KAQKFAOMNZTLHT-OZUDYXHBSA-N	Yeast	96-well plate	Not detected above host organism background levels	Acetyl-CoA	485	5 56
385	Prostaglandin E2	XEYBRNLFEZDVAW-ARSRFYASSA-N	Yeast	96-well plate	Not detected above host organism background levels	Acetyl-CoA	469	5 56
386	Prostaglandin F2alpha	PXGPLTODNUVGFL-YNNPMVKQSA-N	Yeast	96-well plate	Not detected above host organism background levels	Acetyl-CoA	432	5 56
387	Allantoin	POJWUDADGALRAB-SFOWXEAESA-N	Yeast	96-well plate	Not detected above host organism background levels	alpha-D-Ribose 1-phosphate	225	1 70
388	Caffeine	RYYVLZVUVIJVGH-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	alpha-D-Ribose 1-phosphate	293	3 91
389	(S)-N-Methylcoclaurine	BOKVLBSSPUTWLV-INIZCTEOSA-O	Yeast	96-well plate	Not detected above host organism background levels	Chorismate	356	6 74
390	2,3-Dihydroxybenzoate	GLDQAMYCGOIJDV-UHFFFAOYSA-M	Yeast	96-well plate	Not detected above host organism background levels	Chorismate	157	1 57
391	3'-Hydroxy-N-methyl-(S)-coclaurine	DAUPWJBRVZCBQB-AWEZNQCLSA-N	Yeast	96-well plate	Not detected above host organism background levels	Chorismate	393	7 74
392	5,6,7,8-Tetrahydrofolic Acid	MSTNYGQPCMXVAQ-KIYNQFGBSA-N	Yeast	96-well plate	Not detected above host organism background levels	Chorismate	834	0 107
393	Berbamunine	FDABVSXGAMFQQH-XZWHSSHBSA-N	Yeast	96-well plate	Not detected above host organism background levels	Chorismate	902	8 82
394	Enterochelin	SERBHKJMVBATSJ-BZSNNMDCSA-N	Yeast	96-well plate	Not detected above host organism background levels	Chorismate	1,060	4 62
395	Isopropanol	KFZMGEQAYNKOFK-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	Chorismate	11	2 62

396	Vibriobactin	SNUNDXZQRUCXRW-DBIMOCAASA-N	Yeast	96-well plate	Not detected above host organism background levels	Chorismate	1,610	6	100
397	Yersiniabactin	CARQCAHMXWHUIE-NRMQOZBLSA-N	Yeast	96-well plate	Not detected above host organism background levels	Chorismate	914	4	94
398	(-)-Beta-Pinene	WTARULDDTDQWMU-IUCAKERBSA-N	Yeast	96-well plate	Not detected above host organism background levels	geranyl diphosphate (charge:-3)	177	2	51
399	(-)-Menthol	NOOLISFMXDJSKH-KXUCPTDWSA-N	Yeast	96-well plate	Not detected above host organism background levels	geranyl diphosphate (charge:-3)	120	8	58
400	(+)-Comphene	CRPUJAZIXJMDBK-DTWKUNHWSA-N	Yeast	96-well plate	Not detected above host organism background levels	geranyl diphosphate (charge:-3)	177	2	51
401	(+)-Sabinene	NDVASEGYNIMXJL-NXEZZACHSA-N	Yeast	96-well plate	Not detected above host organism background levels	geranyl diphosphate (charge:-3)	179	2	51
402	(+)-Trans-Carveol	BAVONGHXFVOKBV-VHSXEESVSA-N	Yeast	96-well plate	Not detected above host organism background levels	geranyl diphosphate (charge:-3)	191	2	53
403	(6e)-8-Oxogeranial	GRHWFPUCRVCMRY-TXFIJWAUSA-N	Yeast	96-well plate	Not detected above host organism background levels	geranyl diphosphate (charge:-3)	212	4	51
404	1,8-Cineole	WEEGYLXZBRQIMU-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	geranyl diphosphate (charge:-3)	164	2	51
405	3-Carene	BQOFWKZOCNGFEC-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	geranyl diphosphate (charge:-3)	186	2	51
406	7-Deoxyloganate	DSXFHNSGLYXPNG-YDYVGBNJSA-M	Yeast	96-well plate	Not detected above host organism background levels	geranyl diphosphate (charge:-3)	530	7	57
407	7-Deoxyloganetic Acid	DKGYTSKPMLWLEI-FIZOKRMRSA-M	Yeast	96-well plate	Not detected above host organism background levels	geranyl diphosphate (charge:-3)	278	6	53
408	Carvone	ULDHMXUKGWMISQ-VIFPVBQESA-N	Yeast	96-well plate	Not detected above host organism background levels	geranyl diphosphate (charge:-3)	223	4	51
409	Gamma-Terpinene	YKFLAYDHMOASIY-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	geranyl diphosphate (charge:-3)	171	2	51
410	Iridotrial	PFGBAVLSGZLAAY-JYPFKNLRSA-N	Yeast	96-well plate	Not detected above host organism background levels	geranyl diphosphate (charge:-3)	232	5	51
411	Loganin	AMBQHHVBBHTQBF-UOUCRYGSSA-N	Yeast	96-well plate	Not detected above host organism background levels	geranyl diphosphate (charge:-3)	580	9	75
412	P-Cymene	HFPZCAJZSCWRBC-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	geranyl diphosphate (charge:-3)	86	2	63
413	Secologanin	CSKKDSFETGLMSB-NRZPKYKESA-N	Yeast	96-well plate	Not detected above host	geranyl diphosphate (charge:-3)	576	10	76

				organism background levels				
414	Strictosidine	XBAMJZTXGWPTRM-NTXHKPOFSA-N	Yeast	96-well plate	Not detected above host organism background levels	geranyl diphosphate (charge:-3)	887	11 94
415	Terpinolene	MOYAFQVGZZPNRA-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	geranyl diphosphate (charge:-3)	178	2 51
416	Verbenone	DCSCXTJOXBUFGB-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	geranyl diphosphate (charge:-3)	248	3 57
417	(2e,4e,6e)-7-Hydroxy-4-Methylhepta-2,4,6-Trienal	TYQZWSFEJJSSQR-YUHMVAQCSA-N	Yeast	96-well plate	Not detected above host organism background levels	Geranylgeranyl diphosphate	176	6 59
418	2'-Epi-5-Deoxystrigol	QXTUQXRFEBHUBA-MDVYPADHSA-N	Yeast	96-well plate	Not detected above host organism background levels	Geranylgeranyl diphosphate	709	9 60
419	5-Deoxystrigol	QXTUQXRFEBHUBA-DYLOANJQSA-N	Yeast	96-well plate	Not detected above host organism background levels	Geranylgeranyl diphosphate	709	9 59
420	Abscisic Acid	JLIDBLDQVAYHNE-YKALOCIXSA-N	Yeast	96-well plate	Not detected above host organism background levels	Geranylgeranyl diphosphate	494	12 58
421	all-trans-10'-Apo-beta-carotenal	PJEHRCCPERVGEC-FLHUAPOTSA-N	Yeast	96-well plate	Not detected above host organism background levels	Geranylgeranyl diphosphate	771	6 60
422	Alpha-Springene	KFHRKQVLZRJWNB-OBHWEXPVSA-N	Yeast	96-well plate	Not detected above host organism background levels	Geranylgeranyl diphosphate	396	1 59
423	alpha-Tocopherol	GVJHHUAWPYXKBD-IEOSBIPESA-N	Yeast	96-well plate	Not detected above host organism background levels	Geranylgeranyl diphosphate	503	5 86
424	Astaxanthin	MQZIGYBFDRPAKN-UWFIBFSHSA-N	Yeast	96-well plate	Not detected above host organism background levels	Geranylgeranyl diphosphate	1,340	6 59
425	Carlactone	OTIYLZVFQIMLQH-FRGMPSNRSA-N	Yeast	96-well plate	Not detected above host organism background levels	Geranylgeranyl diphosphate	574	6 59
426	Ent-2'-Epi-5ds	QXTUQXRFEBHUBA-OHYDKCKCSA-N	Yeast	96-well plate	Not detected above host organism background levels	Geranylgeranyl diphosphate	709	9 60
427	Ferruginol	QXNWVJOHUAQHLM-AZUAARDMSA-N	Yeast	96-well plate	Not detected above host organism background levels	Geranylgeranyl diphosphate	385	3 55
428	Fusicocca-2,10(14)-Diene	PZSFDLBSQBBRAM-GZRFBZBPSA-N	Yeast	96-well plate	Not detected above host organism background levels	Geranylgeranyl diphosphate	462	1 55
429	Gibberellin A3	IXORZMNAPKEEDV-OBDJNFEBSA-N	Yeast	96-well plate	Not detected above host organism background levels	Geranylgeranyl diphosphate	772	10 59
430	Miltiradiene	BGVUIJDZTQIJIO-AZUAARDMSA-N	Yeast	96-well plate	Not detected above host organism	Geranylgeranyl diphosphate	460	2 55

				background levels				
431	Neoxanthin	PGYAYSRVSAJXTE-CLONMANBSA-N	Yeast	96-well plate	Not detected above host organism background levels	Geranylgeranyl diphosphate	1,380	8	55
432	orobanchol	CDBBMEYPRMUMTR-RZXXLYMMSA-N	Yeast	96-well plate	Not detected above host organism background levels	Geranylgeranyl diphosphate	741	10	59
433	Phylloquinol	BUFJIHPUGZHTHL-NKFFZRIASA-N	Yeast	96-well plate	Not detected above host organism background levels	Geranylgeranyl diphosphate	553	10	94
434	Phytol	BOTWFXYSPFMFNR-PYDDKJGSSA-N	Yeast	96-well plate	Not detected above host organism background levels	Geranylgeranyl diphosphate	255	3	55
435	Picrocrocin	WMHJCSAICLADIN-WYWSWGBSSA-N	Yeast	96-well plate	Not detected above host organism background levels	Geranylgeranyl diphosphate	473	8	62
436	Stemod-13(17)-Ene	GNNRCBBKCVNPSC-VDWQKOAOSA-N	Yeast	96-well plate	Not detected above host organism background levels	Geranylgeranyl diphosphate	445	2	55
437	Strigol	VOFXXOPWCBSPAA-KCNJUGRMSA-N	Yeast	96-well plate	Not detected above host organism background levels	Geranylgeranyl diphosphate	741	10	60
438	Taxa-4(20),11-Dien-5alpha-Ol	QHDGSWAXTYWVOP-ZNWBIBPKSA-N	Yeast	96-well plate	Not detected above host organism background levels	Geranylgeranyl diphosphate	484	3	57
439	Taxadiene	FRJSECSOXKQMOD-HQRMLTQVSA-N	Yeast	96-well plate	Not detected above host organism background levels	Geranylgeranyl diphosphate	462	1	55
440	Violaxanthin	SZCBXWMUOPQSOX-WVJDLNGLSA-N	Yeast	96-well plate	Not detected above host organism background levels	Geranylgeranyl diphosphate	1,270	7	55
441	Zeaxanthin	JKQXZKUSFCKOGQ-QAYBQHTQSA-N	Yeast	96-well plate	Not detected above host organism background levels	Geranylgeranyl diphosphate	1,190	4	58
442	1-Propanol	BDERNNFJNOPAEC-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	Glycerone phosphate	7	3	49
443	3-Hydroxypropionaldehyde	AKXKFZDCRYJKTF-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	Glycerone phosphate	26	1	39
444	Acrylic Acid	NIXOWILDQLNWCW-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	Glycerone phosphate	56	3	45
445	C26 Ceramide	CJROVRTUSFQVMR-GVOPMEMSSA-N	Yeast	96-well plate	Not detected above host organism background levels	L-Serine	651	4	137
446	Deoxyerythronolide B	HQZOLNNEQAKEHT-IBBGRPSASA-N	Yeast	96-well plate	Not detected above host organism background levels	L-Serine	507	3	57
447	L-Cysteine	XUJNEKJLAYXESH-REOHCLBHSA-N	Yeast	96-well plate	Not detected above host organism background levels	L-Serine	75	0	64

42

448	Porphobilinogen	QSHWIQZFGQKFMA-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	L-Serine	270	0	64
449	prodigiosin	WKGQSEFBQTWRPT-UUPRNIFOSA-N	Yeast	96-well plate	Not detected above host organism background levels	L-Serine	612	13	91
450	Taurine	XOAAWQZATWQOTB-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	L-Serine	120	2	58
451	Uroporphyrinogen Iii	HUHWZXWWOFSFKF-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	L-Serine	1,630	0	66
452	Agroclavine	XJOOMMHNYOJWCZ-UKRRQHHQSA-N	Yeast	96-well plate	Not detected above host organism background levels	L-Tryptophan	375	6	90
453	Cyclo[D-His-L-Trp-]	YNHVZXATNGMVNC-LSDHHAIUSA-N	Yeast	96-well plate	Not detected above host organism background levels	L-Tryptophan	499	1	101
454	deoxyviolacein	OJUJNNKCVPCATE-QGOAFFKASA-N	Yeast	96-well plate	Not detected above host organism background levels	L-Tryptophan	680	6	70
455	meleagrine	JNTXZCAYFJHPJL-WSHSOXHMSA-N	Yeast	96-well plate	Not detected above host organism background levels	L-Tryptophan	916	6	114
456	neoxaline	HHLNXXASUKFCCX-LICLKQGHSA-N	Yeast	96-well plate	Not detected above host organism background levels	L-Tryptophan	874	6	115
457	Physostigmine	PIJVFDBKTWXHHD-HIFRSBDPSA-N	Yeast	96-well plate	Not detected above host organism background levels	L-Tryptophan	403	10	84
458	Thaxtomine A	QRDNJYNIEGRRKV-PGRDOPGGSA-N	Yeast	96-well plate	Not detected above host organism background levels	L-Tryptophan	759	4	95
459	1-(4-Hydroxyphenyl)-1-decene-3,5-dione	BWAVVEBSGHWPHL-QUHNVXOFSA-N	Yeast	96-well plate	Not detected above host organism background levels	L-Tyrosine	323	5	70
460	3',4',7-Trihydroxyisoflavone	DDKGKOOLFLYZDL-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	L-Tyrosine	419	8	65
461	Umbelliferone	ORHBXUUXSCNDEV-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	L-Tyrosine	222	4	66
462	(-)-Jasmonic acid	ZNJFBWYDHIGLCU-HWKXXFMVSA-N	Yeast	96-well plate	Not detected above host organism background levels	Malonyl-CoA	268	25	68
463	12-Oxo-Phytodienoic Acid	PMTMAFAPLCGXGK-VOPVTQPFSA-N	Yeast	96-well plate	Not detected above host organism background levels	Malonyl-CoA	382	9	67
464	Aurofusarin	VSWWTKVILIZDGX-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	Malonyl-CoA	1,370	6	137
465	Biotin	YBJHBAHKTGYVGT-ZKWXMUAHSA-N	Yeast	96-well plate	Not detected above host	Malonyl-CoA	298	7	99

					organism background levels				
466	deferoxamine	UBQYURCVBFRUQT-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	Malonyl-CoA	739	8	69
467	Dethiobiotin	AUTOLBMXDDTRRT-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	Malonyl-CoA	243	6	78
468	Emodin	RHMXXJGYXNZAPX-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	Malonyl-CoA	434	3	49
469	Endocrocin	UZOHDKGTYVTYDZ-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	Malonyl-CoA	545	2	49
470	Endocrocin-9-anthrone	XVSLBJJMYVAGDF-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	Malonyl-CoA	477	1	49
471	Hexadecanoic acid	IPCSVZSSVZVIGE-UHFFFAOYSA-N	E. coli	96-well plate	Not detected above host organism background levels	Malonyl-CoA	178	0	67
472	Hexanoic acid	FUZZWVXGSFPDMH-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	Malonyl-CoA	69	2	54
473	hypothemycin	HDZUUVQEDFZKAX-JOLUUEAESA-N	Yeast	96-well plate	Not detected above host organism background levels	Malonyl-CoA	590	4	72
474	Monacolin L	BKZPCUPKVCPRQW-RRYZMYJUSA-N	Yeast	96-well plate	Not detected above host organism background levels	Malonyl-CoA	479	4	68
475	Norsolorinic acid	XIJDBHLQUYAZJI-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	Malonyl-CoA	603	4	51
476	Orsellinic acid	AMKYESDOVDKZKV-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	Malonyl-CoA	180	1	54
477	rubrofusarin	FPNKCZKRICBAKG-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	Malonyl-CoA	432	3	138
478	Trans-2-Hexenal	MBDOYVRWFFCFHM-SNAWJCMRSA-N	Yeast	96-well plate	Not detected above host organism background levels	Malonyl-CoA	65	23	121
479	Cyanin	RDFLLVCQYHQOBU-ZOTFFYTFSA-O	Yeast	96-well plate	Not detected above host organism background levels	Naringenin	899	10	82
480	Hesperetin	AIONOLUJZLIMTK-AWEZNQCLSA-N	Yeast	96-well plate	Not detected above host organism background levels	Naringenin	413	7	78
481	Isorhamnetin	IZQSVPBOUDKVDZ-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	Naringenin	503	8	93
482	Leucocyanidin	SBZWTSHAFILOTE-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism	Naringenin	392	7	66

44

				background levels				
483	Leucodelphinidin	ZEACOKJOQLAYTD-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	Naringenin	408	7 66
484	malvidin	KZMACGJDUUWFCH-UHFFFAOYSA-O	Yeast	96-well plate	Not detected above host organism background levels	Naringenin	406	10 93
485	Peonidin	XFDQJKDGGOEYPI-UHFFFAOYSA-O	Yeast	96-well plate	Not detected above host organism background levels	Naringenin	378	9 84
486	Petunidin	AFOLOMGWVXKIQL-UHFFFAOYSA-O	Yeast	96-well plate	Not detected above host organism background levels	Naringenin	406	9 89
487	Rutin	IKGXIBQEEMLURG-NVPNHPEKSA-N	Yeast	96-well plate	Not detected above host organism background levels	Naringenin	1,020	12 69
488	Sophoraflavanone G	XRYVAQQLDYTHCL-CMJOXMDJSA-N	Yeast	96-well plate	Not detected above host organism background levels	Naringenin	684	8 71
489	1-Butanol	LRHPLDYGYMQRHN-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	Oxaloacetate	13	7 52
490	3-Hydroxybutyrolactone	FUDDLSHBRSNCBV-GSVOUGTGSA-N	Yeast	96-well plate	Not detected above host organism background levels	Oxaloacetate	89	6 54
491	Citric Acid	KRKNYBCHXYNGOX-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	Oxaloacetate	227	0 44
492	desferrioxamine E	NHKCCADZVLTPPO-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	Oxaloacetate	764	8 72
493	D-Glucaric acid	DSLZVSRJTYRBFB-LLEIAEIESA-N	Yeast	96-well plate	Not detected above host organism background levels	Oxaloacetate	202	3 44
494	Ethylene Glycol	LYCAIKOWRPUZTN-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	Oxaloacetate	6	3 46
495	Fumaric Acid	VZCYOOQTPOCHFL-OWOJBTEDSA-N	Yeast	96-well plate	Not detected above host organism background levels	Oxaloacetate	119	0 39
496	Glycolic acid	AEMRFAOFKBGASW-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	Oxaloacetate	40	1 49
497	L-Aspartic acid	CKLJMWTZIZZHCS-REOHCLBHSA-N	Yeast	96-well plate	Not detected above host organism background levels	Oxaloacetate	133	0 53
498	L-Aspartic acid	CKLJMWTZIZZHCS-REOHCLBHSA-N	Yeast	96-well plate	Not detected above host organism background levels	Oxaloacetate	133	0 53
499	Oxalic Acid	MUBZPKHOEPUJKR-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	Oxaloacetate	72	1 32

45

500	2-Phenylacetic Acid	WLJVXDMOQOGPHL-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	Phenylpyruvate (charge:-1)	114	0	58
501	3-phenylpropanol	VAJVDSVGBWFCLW-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	Phenylpyruvate (charge:-1)	75	4	62
502	Atropine	RKUNBYITZUJHSG-PJPHBNEVSA-N	Yeast	96-well plate	Not detected above host organism background levels	Phenylpyruvate (charge:-1)	353	13	155
503	Baicalin	IKIIZLYTISPENI-ZFORQUDYSA-N	Yeast	96-well plate	Not detected above host organism background levels	Phenylpyruvate (charge:-1)	748	8	73
504	Cinnamaldehyde	KJPRLNWUNMBNBZ-QPJJXVBHSA-N	Yeast	96-well plate	Not detected above host organism background levels	Phenylpyruvate (charge:-1)	121	2	113
505	Littorine	FNRXUEYLFZLOEZ-LGGPCSOHSA-N	Yeast	96-well plate	Not detected above host organism background levels	Phenylpyruvate (charge:-1)	353	10	156
506	Penicillin	JGSARLDLIJGVTE-MBNYWOFBSA-N	Yeast	96-well plate	Not detected above host organism background levels	Phenylpyruvate (charge:-1)	530	4	72
507	Styrene	PPBRXRYQALVLMV-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	Phenylpyruvate (charge:-1)	68	2	58
508	(R,R)-Butane-2,3-Diol	OWBTYPJTUOEWEK-QWWZWVQMSA-N	E. coli	96-well plate	Not detected above host organism background levels	Pyruvate	31	2	36
509	Cyclohexanone	JHIVVAPYMSGYDF-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	Pyruvate	68	9	61
510	Diacetyl	QSJXEFYPDANLFS-UHFFFAOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	Pyruvate	72	1	39
511	16-Alpha-Hydroxy-Beta-Amyrin	VLRYIIPJIVGFIV-SIXFPRLASA-N	Yeast	96-well plate	Not detected above host organism background levels	Squalene	825	3	59
512	16-alpha-hydroxy-oleanolic acid	YKOPWPOFWMYZJZ-OKLDCGQTSA-N	Yeast	96-well plate	Not detected above host organism background levels	Squalene	919	4	62
513	Ambrein	BIADSXOKHZFLSN-CQDHYIGWSA-N	Yeast	96-well plate	Not detected above host organism background levels	Squalene	683	2	61
514	Betulinic acid	QGJZLNKBHJESQX-FZFNOLFKSA-N	Yeast	96-well plate	Not detected above host organism background levels	Squalene	861	3	57
515	Brassinolide	IXVMHGVQKLDRKH-KNBKMWSGSA-N	Yeast	96-well plate	Not detected above host organism background levels	Squalene	755	13	98
516	Cucurbitadienol	WSPRAEIJBDUDRX-UPBIWNDXSA-N	Yeast	96-well plate	Not detected above host organism background levels	Squalene	754	2	55
517	Cycloartenol	ONQRKEUAIJMULO-YBXTVTTCSA-N	Yeast	96-well plate	Not detected above host	Squalene	760	2	55

46

				organism background levels				
518	Echinocystic Acid	YKOPWPOFWMYZJZ-PRIAQAIDSA-N	Yeast	96-well plate	Not detected above host organism background levels	Squalene	919	4	61
519	Ecliptasaponin A	WYDPEADEZMZKNM-ZBKPBKBGSA-N	Yeast	96-well plate	Not detected above host organism background levels	Squalene	1,210	5	62
520	Ginsenoside Rg3	RWXIFXNRCLMQCD-CZIWJLDFSA-N	Yeast	96-well plate	Not detected above host organism background levels	Squalene	1,370	5	63
521	Ginsenoside Rh2	CKUVNOCSBYYHIS-IRFFNABBSA-N	Yeast	96-well plate	Not detected above host organism background levels	Squalene	1,070	4	59
522	Gypsogenin	QMHCWDVPABYZMC-MYPRUECHSA-N	Yeast	96-well plate	Not detected above host organism background levels	Squalene	936	4	54
523	Protopanaxatriol	SHCBCKBYTHZQGZ-DLHMIPLTSA-N	Yeast	96-well plate	Not detected above host organism background levels	Squalene	817	4	55
524	Ursolic Acid	WCGUUGGRBIKTOS-GPOJBZKASA-N	Yeast	96-well plate	Not detected above host organism background levels	Squalene	874	3	55
525	Vitamin D3	QYSXJUFSXHHAJI-YRZJJWOYSA-N	Yeast	96-well plate	Not detected above host organism background levels	Squalene	610	4	80
526	Anditomin	HWLYZRWDCDSFFO-VIJQBOBYSA-N	Yeast	96-well plate	Not detected above host organism background levels	trans,trans-Farnesyl diphosphate	976	17	76
527	Capsidiol	BXXSHQYDJWZXPB-OKNSCYNVSA-N	Yeast	96-well plate	Not detected above host organism background levels	trans,trans-Farnesyl diphosphate	358	2	54
528	Germacrene A Acid	IBJVPIJUFFVDBS-JBMXZMKISA-N	Yeast	96-well plate	Not detected above host organism background levels	trans,trans-Farnesyl diphosphate	361	3	56
529	Germacrene D-4-Ol	RHCTXHCNRLCYBN-WSLVBCCMSA-N	Yeast	96-well plate	Not detected above host organism background levels	trans,trans-Farnesyl diphosphate	275	1	59
530	Heme A	ZGGYGTCPXNDTRV-YRJWGCQVSA-L	Yeast	96-well plate	Not detected above host organism background levels	trans,trans-Farnesyl diphosphate	2,130	0	84