Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Systems Biology

Retrosynthetic design of metabolic pathways to chemicals not found in nature

Geng-Min Lin[1], Robert Warden-Rothman[1] and Christopher A. Voigt

Abstract

Biology produces a universe of chemicals whose precision and complexity is the envy of chemists. Over the last 30 years, the expansive field of metabolic engineering has many successes in optimizing the overproduction of metabolites of industrial interest, including moving natural product pathways to production hosts (*e.g.*, plants to yeast). However, there are stunningly few examples where enzymes are artificially combined to make a chemical that is not found somewhere in nature. Here, we review these efforts and discuss the challenges limiting the construction of such pathways. An analogy is made to the retrosynthesis problem solved in chemistry using algorithmic approaches, recently harnessing artificial intelligence, noting key differences in the needs of the optimization problem. When these issues are addressed, we see a future where chemistry and biology are intertwined in reaction networks that draw on the power of both to build currently unobtainable molecules across consumer, industrial, and defense applications.

Addresses

Synthetic Biology Center, Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

Corresponding author: Voigt, Christopher A (cavoigt@gmail.com)

[1] Equally contributing authors.

Current Opinion in Systems Biology 2019, **14**:82–107

This reviews comes from a themed issue on **Synthetic biology**

Edited by **Jeff Tabor** and **Julius Lucks**

For a complete overview see the Issue and the Editorial

Available online 18 April 2019

https://doi.org/10.1016/j.coisb.2019.04.004

Keywords

Synthetic biology, Systems biology, Artificial intelligence, Machine learning.

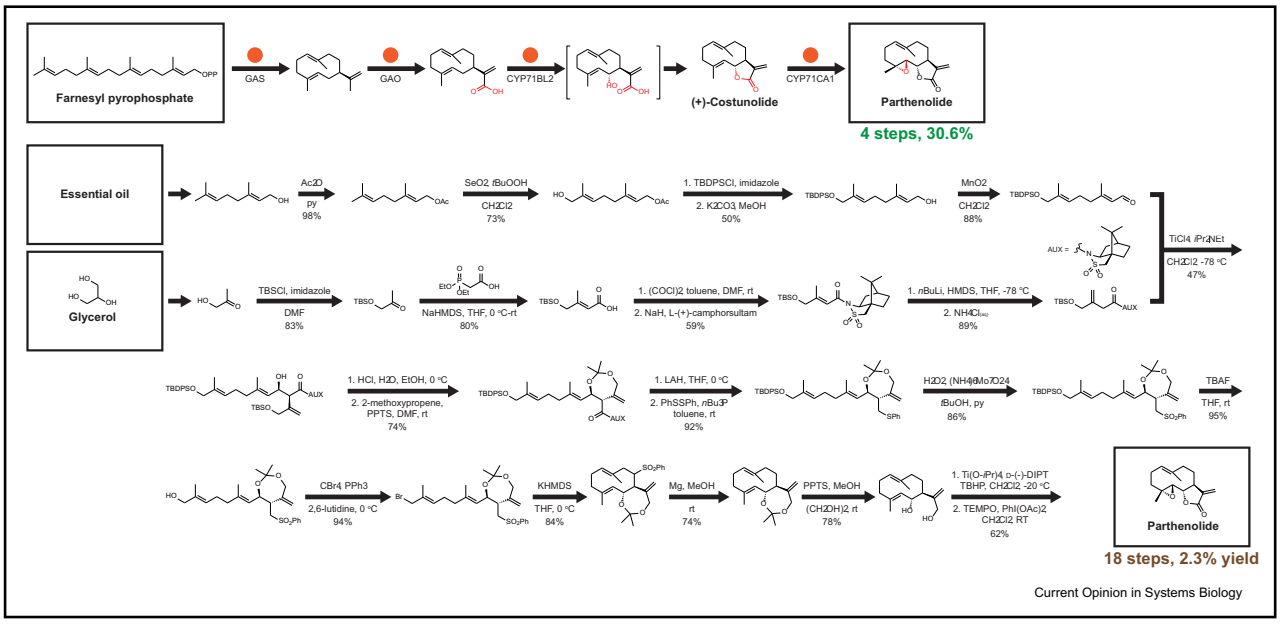
Introduction

When posed with a new molecule to build, organic chemists have to work backwards. Known as 'retrosynthesis,' they identify the series of chemical transformations that can construct the target from simpler chemical building blocks derived from petroleum or other sources [1]. Chemicals that are large and complex with many functional groups and stereocenters have required Herculean efforts to build; for example, halichondrin B has 32 stereocenters (4 billion isomers) and requires a sprawling total synthesis whose longest linear path is 47 reactions [2]. Solutions have been found to incredibly challenging retrosynthesis problems, many of which are recorded in the tomes Classics in Total Synthesis I, II, and III [3–5]. Human creativity has been the power behind these solutions, but as the literature size has grown to an unmanageable size, computer algorithms have been developed to guide the identification of reaction paths [**6].

Cells are the envy of chemists as they are able to build complex chemicals at high yields under ambient conditions. They excel in dictating patterns of stereochemistry, and their products are impossibly functionalized. In fact, many of the most lauded retrosynthesis efforts are to rebuild the chemicals made by biology, for example, hard-to-source pharmaceuticals, which can be built with fewer steps biosynthetically (Figure 1). With a view of the complex chemicals built by the natural world, it is clear that it would be revolutionary to be able to harness these processes to build unnatural molecules of such complexity by design.

Reaching this vision has been inhibited by several challenges. Only a fraction of the enzymes sequenced have been characterized. Their high specificity makes it more difficult to mix and match them between pathways as part of retrosynthetic effort. Before the emergence of DNA synthesis, it was difficult to come by the genetic material for enzymes described in the literature [7]. Physically combining enzyme-encoding genes to build a pathway has only recently been made possible with genetic part libraries to control expression, DNA assembly to put the enzymes together, 'Foundries' to build and debug designs, and methods to work with nonmodel organisms [8–16]. Directed evolution has been able to expand enzyme specificity to more substrates and catalyze new reactions, including those not known to occur in nature [17,18]. Artificial intelligence is increasingly being applied to the design process.

Figure 1



Comparison of metabolic and chemical routes to parthenolide [339,340]. The pathway has been identified and transferred from its native organism (*Tanacetum parthenium*) to yeast, and the theoretical yield of the biosynthetic route is shown (0.306 g/g glucose).

The rise of the bioeconomy will see biochemistry merge with organic chemistry as the means by which novel chemicals are discovered and produced. This manuscript reviews the first steps of this process towards the automation of retrosynthetic design in biochemistry. Like its counterpart in chemistry, the goal of metabolic retrosynthetic design is to start with a desired chemical and then determine the enzymatic transformations required to build it from metabolites made by the cell. First, we review the metabolic retrosynthesis projects completed to date. As in chemistry, they have relied on human expertise and creativity, and the largest projects have been to rebuild natural products in a production strain (*e.g.*, move a plant pathway to yeast). Second, we will describe efforts to mine enzymes and evolve their specificity and ability to perform reactions not found in nature. Finally, we will review recent efforts to develop algorithms where a user can specify a chemical structure and the software identifies the combination of enzymes required to build it in a cell.

Classics in total retrobiosynthesis

Moving natural product pathways from plants to production hosts, such as yeast, makes up the most complex metabolic engineering projects to date. All the biosynthetic enzymes are present in the plant, and in some cases, they can all be moved into the new host together [19–22]. This is not always possible as an enzyme may be unknown or it is nonfunctional in the new host due to misfolding, localization, missing cofactors, precursor supply, feedback inhibition, or differences in the intracellular conditions. This is solved by replacing portions of the pathway sourced from different organisms. Figure 2a shows published pathways that are not entirely sourced from a single organism and require more than two enzymatic steps from the nearest cellular metabolite [**23–56].

The route to noscapine, an anticancer alkaloid derived from the opium poppy, is the most complex example to date [**23]. A pathway was constructed in yeast to make noscapine from tyrosine, requiring 18 enzymatic steps, of which only 13 involve enzymes from the opium poppy (*Papever somniferum*). Of the remaining five, one is from the brown rat (*Rattus novegicus*), one from bacteria (*Pseudomonas putida*), and three from other plants (*Coptis japonica* and *Eschscholzia californica*) [**23,*57–59]. Another example is the reconstruction of the tetrahydrocannabinol (THC) pathway from *Cannabis sativa* in yeast [39,60–64]. Only 5 enzymes were from the native plant, with the remaining 6 sourced from four diverse bacterial species. Both the noscapine and THC projects required protein engineering to alter the enzymes to eliminate feedback inhibition and alter localization tags. Broad enzyme substrate specificity was recognized as being critical in combining enzymes from diverse sources to form a new pathway.

Xenobiotic molecules are those not known to occur in nature. Retrobiosynthetic design has been applied to build these compounds by combining a series of enzymes that synthesize it from a cellular metabolite

[**24,26,**28,41–47,65–71]. To our knowledge, only a few such pathways have been constructed that combine three or more enzymatic steps (Figure 2b) [**28,**44,**45]. The construction of these pathways is challenging because, by definition, enzymes have to be found to act on chemicals that are different than their native substrates. As a result, the molecules produced to date are much simpler than those obtained by rebuilding natural pathways and do not approach the complexity known to be possible with biology. The longest published retrobiosynthetic pathway is from the metabolite *a*-ketoglutarate to 6-aminoacprotic acid (6-ACA), which is a precursor for nylon-6 [*72]. Constructed in *Escherichia coli*, it combines enzymes from archea (*Methanococcus aeolicus*) and bacteria (*Azobacter vinlandii*, *Lactococcus lactis*, and *Vibrio fluvialis*) [26,*72–76]. To find these, a panel of 3–15 enzymes had to be screened for each catalytic step, and the top enzymes were noted to have broad substrate specificity [*72,73].

Metabolic pathways can also be introduced into cells to modify a chemical added to the media that is not a natural metabolite. Retrosynthetic design can be applied to identify enzymes from different sources to convert the chemical to a desired product (Figure 2c) [54–56]. Both azelaic acid and carvolactone are found in nature but are not part of the *E. coli* metabolome. An example of a xenobiotic compound is *p*-bromophenylglycine, which was produced as part of an effort to create green routes to α-amino acids that avoid cyanide-based chemistry [56]. This pathway combines four heterogenous enzymes from diverse bacteria (*Pseudomonas* sp. VLB120, *Sphingomonas* sp. HXN-200, *Pseudomonas putida*, and *Streptomyces coelicolor*) with a step performed by a native *E. coli* enzyme.

Mining the world's enzymes

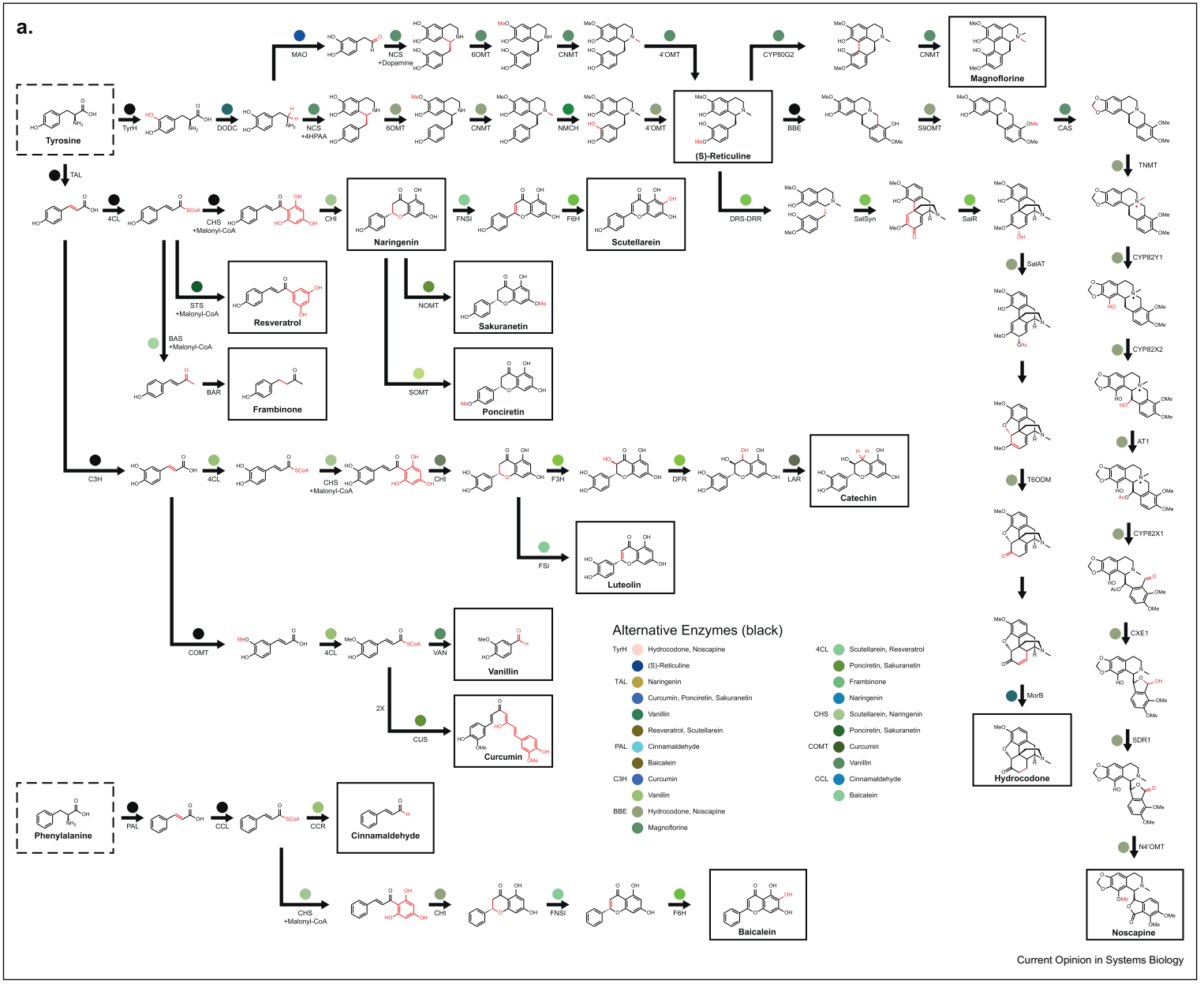
Metabolic retrosynthesis projects require having access to enzymes that will perform the necessary catalytic steps to convert a metabolite into the desired product. An enzyme's specificity is defined as the range of substrates on which it will perform its function. Some enzymes, especially those in central metabolism, are highly specific whereas others are promiscuous. When building a pathway, the specificity profile is crucial information to select an enzyme to function on an intermediate that is not its natural substrate. Making this choice is complicated by the relatively few well-characterized enzymes, particularly with respect to specificity information. However, the convergence of new technologies suggests that we are reaching an inflection point and a deluge of enzyme data is forthcoming.

The number of enzymes represented in earth's biodiversity is extraordinary, and only a tiny fraction has been experimentally characterized. DNA sequencing and environmental metagenomics are starting to illustrate this potential (Figure 3a). Of the estimated 10^{31} nucleotides on earth, 10^{12} have been sequenced and are available in the NCBI database (as of 2019) [77]. It is estimated that the database holds 40 million unique enzymes, a number that doubles every 24 months [78–*82]. Experimentally keeping up with this seems overwhelming, but the convergence of informatics, DNA construction, automation, and screening technologies are increasingly exploiting the information in databases as a resource.

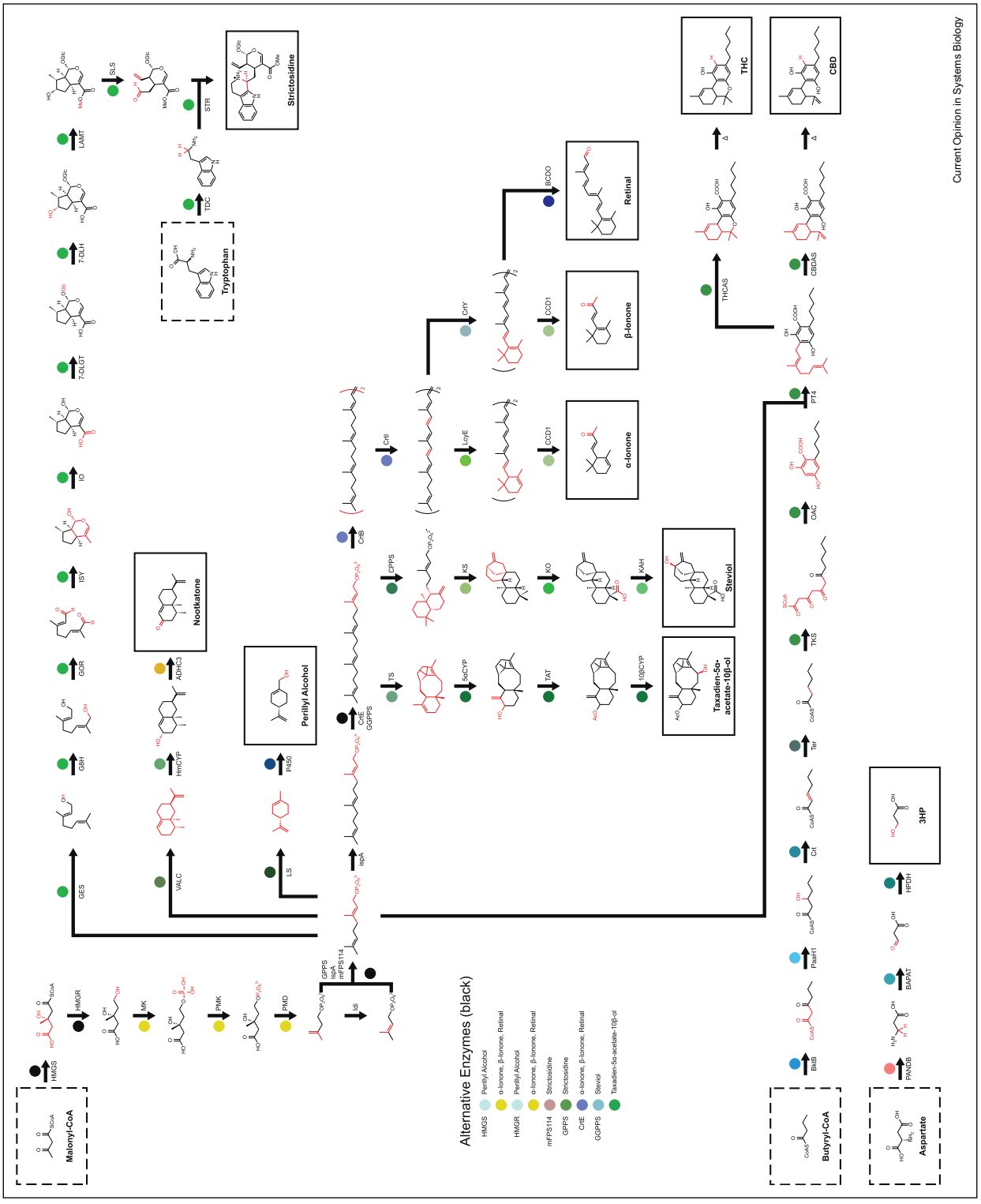
Part mining refers to the now common practice of converting sequence information in a database to physical genes and then screening for a desired function. Figure 3b shows enzyme mining efforts described in the literature, and many more such screens have been performed at companies to populate internal nonpublic databases. Bioinformatic tools to select genes have been getting more sophisticated. Originally, only sequence similarity to a target enzyme guided the search using algorithms such as the Basic Local Alignment Search Tool (BLAST) [78,83–95]. Metabolic databases have compiled reaction and pathway information (Kyoto Encyclopedia of Genes and Genomes (KEGG) and MetaCyc) [96–98]. BRENDA contains manually curated biochemical data for enzymes, including mutants, and substrate profiles [99]. It is increasingly common for genes to be sequenced in the context of genomes (Figure 1a), and local genetic context can be used to help predict its substrate [100–**102]. If one is trying to mine enzymes with diverse activities or substrate profiles, sequence similarity networks can help guide diversity into unexplored regions [103–106]. Homology modeling and computed ligand docking, while notoriously inaccurate in making specific predictions, are still useful to downselect sequences for screening [**104–106].

Once a set of genes is identified, physically building the DNA is straightforward, given the cost, accuracy, and throughput improvements achieved by synthesis companies. Even a low-budget project can now build 100s of enzymes, and while building all 255,463 sequenced P450 enzymes in NCBI is currently cost limited (~$30 million), the continuous decline in price may soon even make that possible [78,107,108]. Screening even a small set of genes can be time consuming, particularly when the chemistry is not conducive to a high-throughput assay. This has been addressed recently through robotic automation, miniaturization of the reaction vesicle using droplets (10^8 reactions per day), and microcapillaries (10^6 per chip) [*105,**109–114]. It is more challenging to obtain the substrate profiles for a large number of enzymes, as opposed to their activity against a single

Figure 2

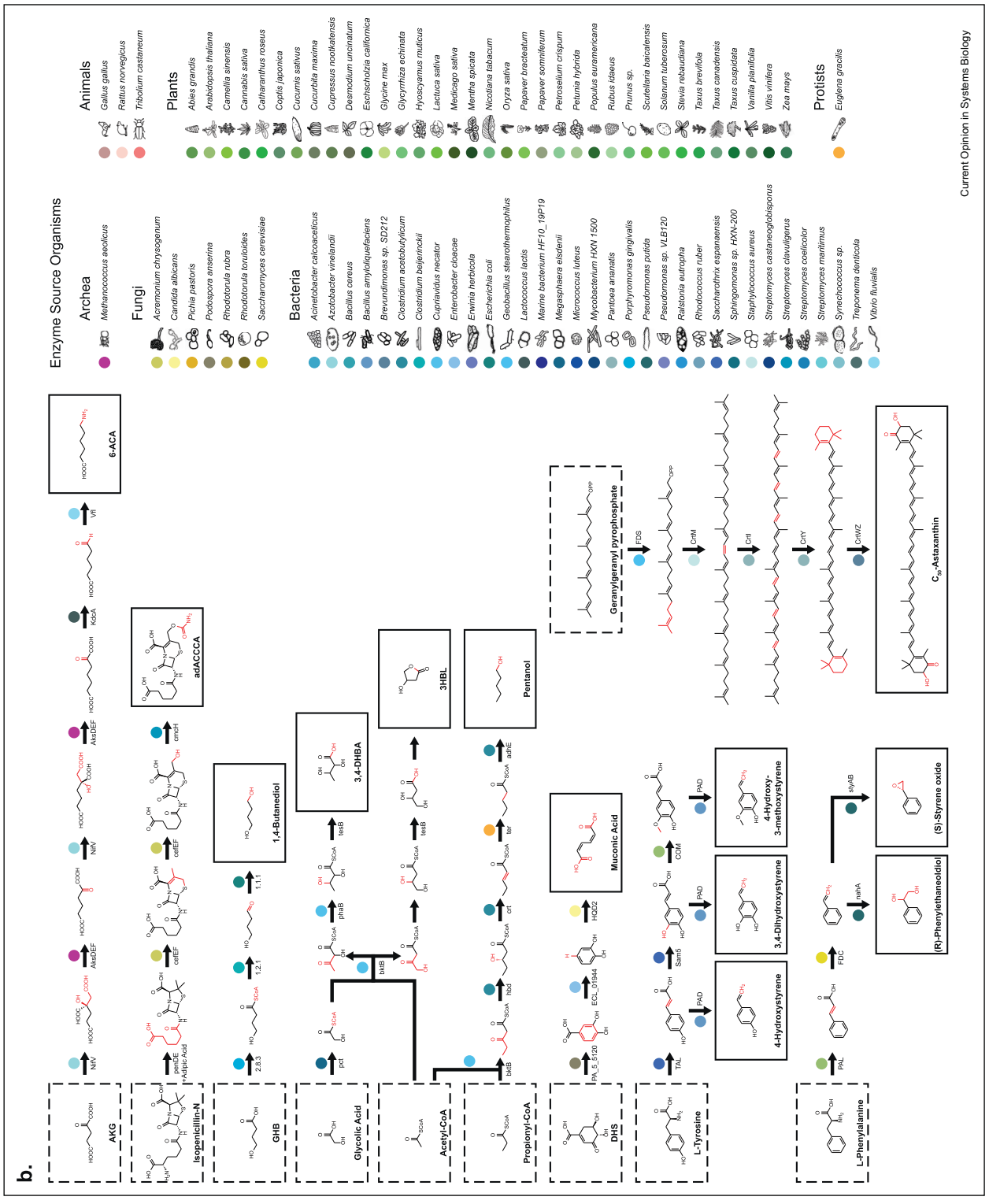


Classics in total retrobiosynthesis are shown for the production of **(a)** natural chemicals in a heterologous host, **(b)** xenobiotic compounds, and **(c)** the conversion of exogenous chemicals to a product [**23–40,42,**44,46,48–56,68–71]. Pathways are shown for 42 products made from 15 precursor metabolites or exogenous chemicals (black boxes). The pathways are based on enzymes from 80 species (colored dots above reaction arrows and legend). The pathways are compiled from multiple projects, and if different enzymes were used to perform the same reaction across these projects, these are shown as a black dot (insets expand to show all enzyme sources). Enzymes native to the chassis organism are named but without a colored dot. Spontaneous reactions have no labels or dots. This figure is intended to be comprehensive for all the pathways we could identify in the literature. Some famous metabolic engineering examples are excluded because they involve the production of native metabolites (e.g., isobutanol) or require less than two steps from a native metabolite (e.g., propanediol) or all the enzymatic steps from a native metabolite are sourced from a single organism (e.g., artemisinin) [65,86,341–344].



(Continued).

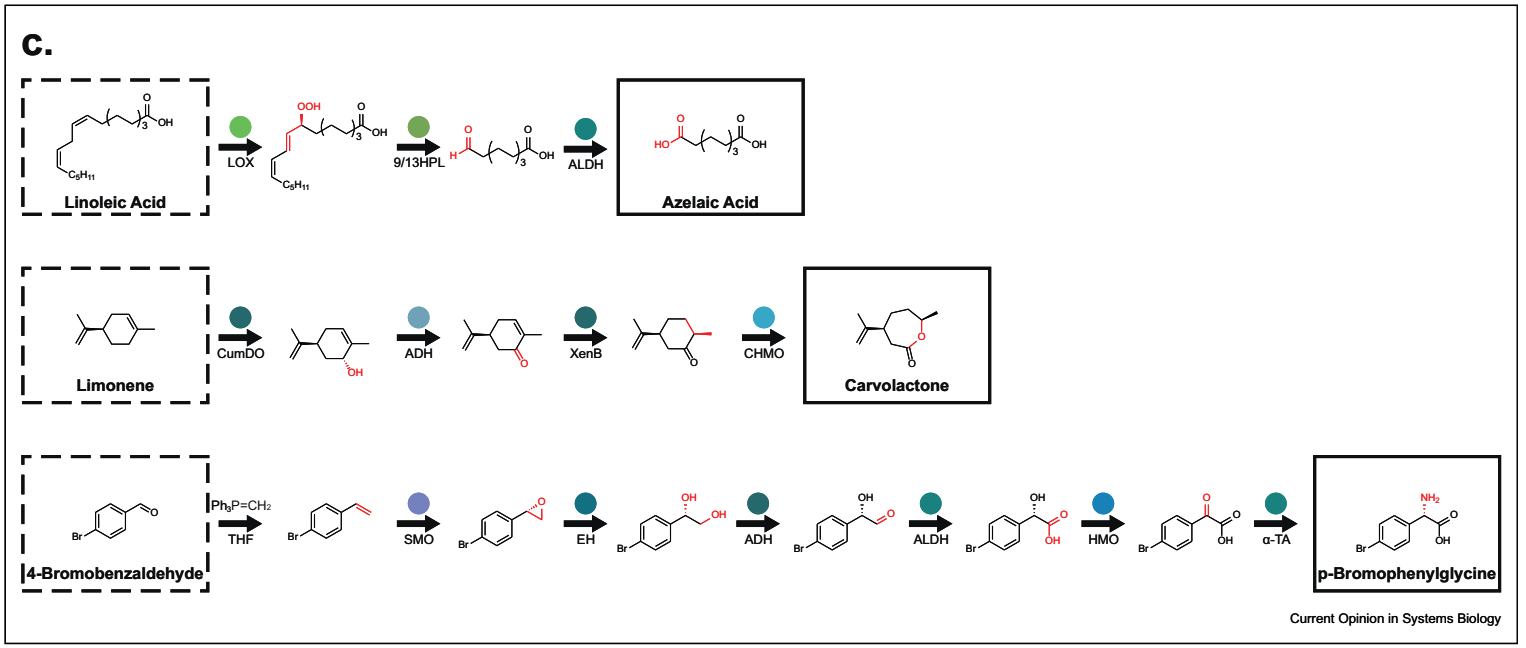
Figure 2



(Continued).

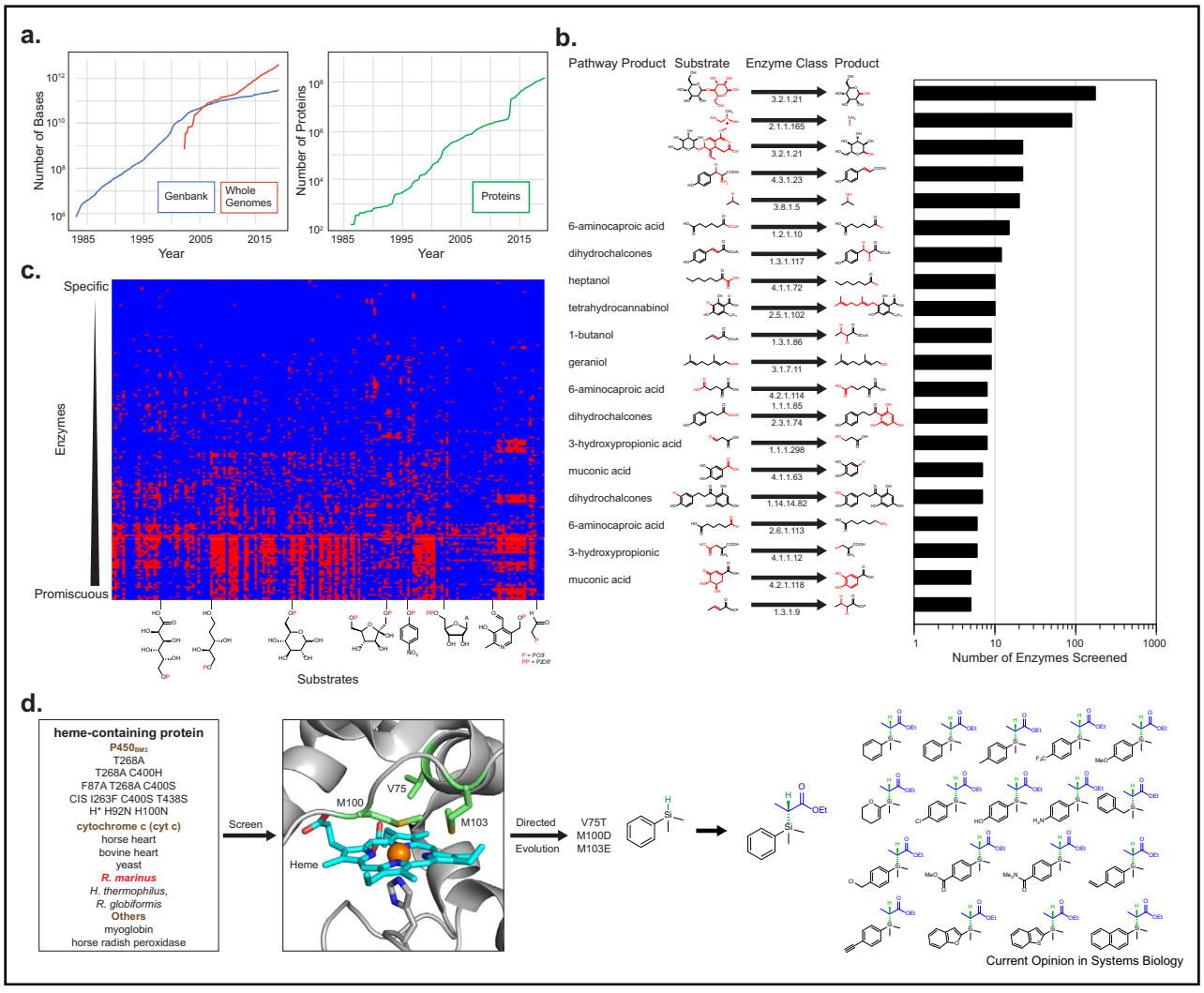
Figure 2

Figure 2



(Continued).

Figure 3



Enzyme mining efforts. **(a)** Number of bases in NCBI Genbank database and proteins in NCBI RefSeq database [81,345] **(b)** Enzyme mining efforts from the literature that involve five or more candidates obtained from a sequence database using DNA synthesis [27,**28,39,*72,87–89,92–95,**104,*105,346]. If the mining effort was part of a metabolic engineering project, the target molecule is listed. The EC number of each enzyme is shown along with the reaction for which it was screened. **(c)** A massive enzyme specificity screen is shown for 217 haloalkanoate dehalogenases against 167 substrates (of which 8 representative examples are shown on the x-axis) [**126]. Red dots show the substrates that each enzyme will catalyze. **(d)** A combined enzyme mining–directed evolution experiment is shown to make an enzyme that catalyzes the enantioselective carbene insertion into Si-H bond [**122]. The resulting enzyme is promiscuous, and the varied products are shown.

substrate. When the reaction is coupled (e.g., NAD^+ is produced), then this can form the basis for a high-throughput screen that is substrate agnostic [115–118]. Substrates can also be spotted onto a microarray and a specificity profile obtained in a single experiment, although this will not be compatible with all chemistries [119]. A more generalizable approach is to use metabolomics to multiplex many substrates (10^3) in a single reaction [115,**120].

Enzyme promiscuity

Promiscuous enzymes are often preferred when building a new biosynthetic pathway or when initiating an enzyme engineering project for a new substrate target [26,**28,121,**122]. Even a small amount of activity can provide the 'hook' that can then be optimized via directed evolution [112,123–125]. The degree of promiscuity can widely vary even within the same enzyme class. In one extreme case, a dehalogenase was active against 143 of 169 substrates tested, whereas 53/217 of the enzymes were active on <5 substrates (Figure 3c) [**126]. When evolving an enzyme to a new target, it is often observed that intermediate mutants increase in promiscuity before increasing activity and specificity in later rounds [123,127–131]. Even a single mutation can switch an enzyme from being specific to promiscuous. A

goal could be to create a 'universal' set of enzyme functions with representatives of each possible reaction type. For part mining, bioinformatic tools may be able to make mutations prior to synthesis that are predicted to increase the promiscuity of the target, for example, by creating a more spacious ligand-binding pocket [80,**104,*105,130–144].

Enzymes also exhibit catalytic promiscuity, where the catalytic residues perform alternative reaction chemistries [18,136,145–149]. Enzymes can be evolved to perform functions not found in nature and difficult to perform chemically, such as stereospecific carbene insertion into carbon–carbon [**150], silicon–hydrogen [**122], or boron–hydrogen [**151] bonds. The parent protein can be surprising: an enzyme that produced enantiopure organosilicon chemicals was found by making three mutations to cytochrome *c*, an electron transfer protein that does not perform a catalytic function in nature (Figure 3d) [**122]. This suggests that organizing part mining efforts by cofactor (e.g., heme-containing) or potential chemistries may aid screening efforts. Computational protein design has also been used to initiate new catalytic functions into a scaffold, including a retro-aldolase, Kemp eliminase, and Diels-Alderase [152–155]. Even though the initial activities are usually low, the computational methods can provide fodder for directed evolution optimization [**109].

Promiscuity is not a problem when purified enzymes are used as catalysts for a chemical process; in fact, it is an advantage as the same enzyme could be integrated into the processing of multiple chemical reactions [121,**122,138]. However, it can pose problems when trying to combine enzymes to build a biosynthetic pathway. A promiscuous enzyme is more likely to act unintentionally on natural cellular metabolites. The resulting products draw resources away from the cell, could be toxic, and these reactions could use up energy and redox resources (*e.g.*, ATP and NADPH) [156]. They can also generate multiple products, complicating purification [**104]. This creates a tension in the discovery process, where promiscuity makes it easier to find an enzyme that acts on a new target but harder to incorporate into an *in vivo* biosynthetic pathway. Computational tools and screening strategies have been developed to aid this process [**104,157].

Chemistry: towards retrosynthesis by artificial intelligence

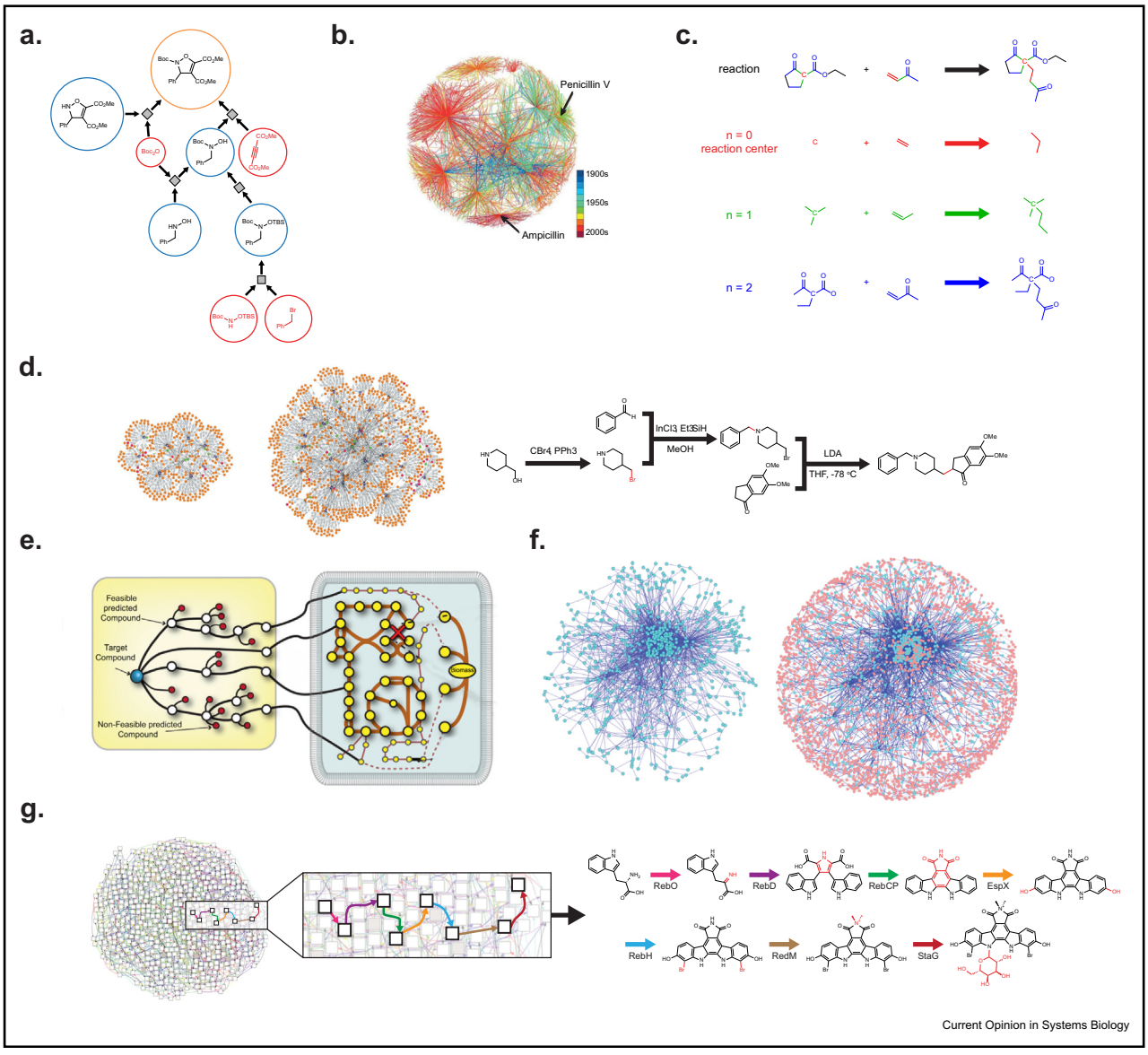
Since the 1960s, the potential for computer algorithms to augment the retrosynthetic decision-making process has been recognized [158,159]. The fundamental challenge of retrosynthesis, both for a human and a computer, is that there is a combinatorial explosion of possible reactions as one iteratively steps backwards towards simpler compounds. One can imagine this as a graph, where each node is a chemical and the edges are reactions (Figure 4a). The goal of the algorithm is to identify a path from the target molecule to precursors that are commercially available [**6]. This search is nontrivial: for the halichondrin B example involving 47 steps, assuming an average of 100 potential reactions per intermediate [**160], there are $\sim 10^{94}$ paths. Retrosynthetic software packages differ in how they build this space using reaction rules, search the space, and weigh individual steps and complete pathways (Table 1). Artificial intelligence is increasingly being applied to problems in organic chemistry, including all aspects of retrosynthesis [**161–170].

As the chemical literature has grown, it has become increasingly challenging to manage the decision-making process and make researchers aware of the possibilities. To address this, a 'Network of Organic Chemistry' has been constructed on the basis of every reaction published in the literature (maintained in the Reaxys database) [163,164,171–175]. As of 2019, the network has 35 million chemicals (1000-fold more than the human metabolome) [**6]. This graph can be used for retrosynthesis by finding the node corresponding to the target molecule and then using a routing algorithm to identify a path to commercially available substrates (Figure 4b) [**6,176–**178]. Each path is feasible because it corresponds to a series of experimentally validated reactions. However, this approach is limited by only being able to predict how to build molecules present in the network and the reactions are sparse, thus limiting the possibility of discovering creative or overlooked possibilities.

The alternative is to condense the reaction information into 'rules' that represent moves that can be made in the networks. Rules represent generic reactions that can be applied to obtain the target chemical when certain functional groups are present, to generate precursor(s) one-step back. The rules may include the atoms surrounding the reaction centers (chemical environment, Figure 4c) and incompatible functional groups that should be avoided. They represent an informed prediction that allows them to be applied to chemicals for which there is not an exact match in the literature. There is a trade-off where too little detail in the rule leads to unrealistic chemistry but too much reverts to the specific reactions (those described by the Network of Organic Chemistry) [163,175]. Each algorithm has a different approach to deal with this trade-off in an attempt to build accurate yet broadly applicable rules.

Early retrosynthesis algorithms were based on databases of rules that required meticulous data entry by experts [**6,163,168,169]. For example, the Chematica software uses a set of ~60,000 rules [**6]. The exponential growth of the literature has made human curation no longer feasible. ARChem Route Designer automatically extracts rules from reaction databases, but its focus on atoms close to the reaction center leads to errors because

Figure 4



Retrosynthetic design algorithms for organic chemistry and metabolic engineering. **(a)** A simple retrosynthetic search is shown from a starting compound (orange) to the available compounds (red). Adapted from Ref [**161]. **(b)** A small subnetwork (~0.1%) of Network of Organic Chemistry is shown, where the nodes are chemicals and edges are reactions [**6]. Edge colors correspond to the year when the reaction was reported, showing the growth of chemical knowledge. **(c)** An example of a generic reaction rule based on the distance (n) to the reaction center. **(d)** A demonstration of a retrosynthetic search is shown by Chematica after 8 (left) and 35 (right) iterations after defining donepezil as the target compound. A top-ranked pathway is shown. Adapted from Ref [**6]. **(e)** A scheme is shown of the retrosynthetic search for a path from a target compound to the nearest metabolite produced by the cell. The image is taken from a description of GEM-Path [**45]. **(f)** The *E. coli* metabolic network (blue, left) is shown embedded in a greater network of all hypothetic metabolic reactions on the basis of reaction rules (right) [208]. **(g)** A directed graph describing potential paths from L-tryptophan to 540 bisindoles. An example of retrobiosynthetic plan for 1,11-dibromo-2,10-dihydroxy-6,6-dimethyl-12-glucosyl-indolo[2,3-a]pyrrolo[3,4-c]carbazole-6-ium-5,7-dione is highlighted. Adapted from Ref. [10].

of long-distance effects (*e.g.*, the acceleration of a leaving group by an allosteric functional group) or conflicts because of multiple reactive groups [175,179]. Increasing the number of atoms and functional groups further from the reaction center leads to more realistic reaction paths [180]. Alternative approaches have been used to predict the transformation of products to reactants that do not require explicit rules. The reaction database can be searched to determine whether structurally similar molecules have been synthesized in order to suggest a retrosynthetic step [**181]. A neural network has been trained on the USPTO database to map a product (encoded as a SMILES (Simplified Molecular Input Line Entry System) string) to a reactant (SMILES string) [*166].

Table 1

Software for retrosynthesis.

Name	Reactions	Reaction source	Ranking factors	Website	Ref
Organic chemistry					
ChemPlanner (ARChem Route Designer)	Reaction signatures (rules)[a]	MOS [347] Beilstein Crossfire[b] [173]	Scoring function	https://www.cas.org/products/scifinder-n/chemplanner	[179]
IC*SYNTH*	Reaction signatures (rules)[c]	SPRESI[web] [348]	Scoring function	https://www.infochem.de/synthesis/ic-synth	[180]
Synthia™ (Chematica)	Bipartite graph/hand-curated rules[d]	Reaxys [173]	Cost/scoring function	https://www.sigmaaldrich.com/chemistry/chemical-synthesis/synthesis-software.html	[**6,**160,**184]
Metabolism					
BNICE	Bond-electron matrices (rules)[e]	KEGG [96]	Reaction ΔGs	http://lcsb-databases.epfl.ch/pathways/atlas/	[195]
enviPath	SMIRKS [349] (rules)[f]	EAWAG-BBD [350]	Reaction 'likelihood'	https://envipath.org/	[214]
NovoStoic	Reaction signatures (rules)[a]	MetRxn [351]	Reaction ΔGs Profit margin	https://github.com/maranasgroup/rePrime_novoStoic	[**215]
PathPred	RDM patterns[g]	KEGG [96], BRENDA [99]	Structural similarity	https://www.genome.jp/tools/pathpred/	[212]
RetroPath2.0	SMIRKS (rules)[f]	MetaNetX [352]	Sequence availability Flux	https://www.myexperiment.org/workflows/4987.html	[**213]
XTMS	Reaction signatures (rules)[a]	MetaCyc [97], KEGG [96]	Promiscuity Path length Unknown steps Toxicity reaction ΔG	http://xtms.issb.genopole.fr/	[209]

[a] Reaction center and relevant neighboring functional groups.

[b] MOS (Methods in Organic Synthesis) is now Synthetic Reaction Updates, and Beilstein Crossfire is now Reaxys.

[c] Reaction center, first-degree, and second-degree neighboring atoms.

[d] Generic reaction description, compatible functional groups, required protecting groups, and reaction condition.

[e] Matrices recording bond breaking and formation around the reaction center.

[f] Text-based reaction pattern matching around the reaction center.

[g] Vectors encoding atom types gained and lost in a reaction.

Once the rule set is defined, branching algorithms start from the target chemical and search the space [**6,182,183]. The search ends when it finds chemicals that are commercially available (*e.g.*, the ~200,000 compounds available in the Sigma-Aldrich catalog) [**160]. Enumerating all possible paths is computationally impossible, so the algorithms use a metric to determine which rules should be pursued at each step, thus stopping the algorithm from pursuing unpromising branches (Figure 4d). All software packages (Table 1) use metrics designed to avoid reactions involving strained intermediates, unlikely structures, or those that are nonselective [**6,**160,179,180]. Rules can also be eliminated where chemicals have functional groups that are sensitive to the reaction conditions [182]. Recently, an AI approach was developed that combines a Monte Carlo tree search with three neural networks to select the best candidate reactions for expansion [**161]. These neural networks are trained to evaluate only the reaction center atoms, the reaction center and first-degree neighboring atoms, and predict feasibility on the basis of augmented negative reaction data from the literature. In a provocative study, Molga et al. showed that Chematica could be applied to identify reaction paths that circumvent patent portfolios by locking not-to-be-altered bonds or motifs during the search [**184].

The algorithms typically find many putative paths, which are then ranked by an objective function. This can bias the solution towards paths that are low cost, reduced labor, environmentally friendly, have the fewest steps, go through the most 'popular' intermediates, avoid regulated compounds and toxic intermediates, or whatever a user defines as important [**6,185].

Despite development over decades, there have only recently been published success stories regarding the automated discovery of new and elegant reaction paths to difficult-to-synthesize compounds. *ICS*SYNTH was shown to be able to suggest an unusual path to the pharmaceutical oxaspiroketone that was experimentally verified to improve yield [180]. In a stunning effort, retrosynthetic paths were predicted by Chematica and experimentally validated for eight high-value and biologically active compounds (BRD7/9 inhibitor, α-hydroxyetizolem, ATR kinase inhibitor, an inhibitor of acute-myeloid-leukemia cell proliferation, *(S)*-4-hydroxyoxyduloxetine, 5β/6β-hydroxylurasidone, dronedarone, and the natural product engelheptanoxide C) [**160,186]. These solutions reduced the number of steps, generated significant yield improvements, identified unique and simpler routes, used alternative source materials, lowered costs, and reduced synthesis time.

Biochemistry: finding a retrosynthetic route through metabolism

In metabolic engineering, the goal of retrosynthesis is to identify a path of enzymatic reactions linking a desired molecule to a cellular metabolite precursor [187,188]. While articulated similarly to the organic chemistry retrosynthesis problem, there are some important differences. Foremost, all the biochemical reactions in a cell occur concurrently as a system, rather than as a series of discrete steps, and a heterologous pathway operates in this context [189]. It is the ultimate '1-pot' reaction where all the reaction conditions (solvent, temperature, and so on) are essentially identical. Enzyme promiscuity is common; for example, 37% of native *E. coli* enzymes have more than one substrate [190]. This leads to a highly interconnected network where all the reactions may not be known. The carbon from the feedstock (*e.g.*, sugar) can be diverted down different paths and away from the desired product, thus reducing the yield (mass product/mass feedstock), a key metric for economic success [187]. There are also many alternative paths by which one metabolite can be converted into another, each of which could be targeted by an engineer to improve flux.

Software has been developed to facilitate metabolic retrosynthesis (Table 1) [187,188,191–194]. When a user defines a target molecule, the algorithms search for a reaction path connecting it to a list of acceptable starting metabolites (Figure 4e). This search is performed on a metabolic network, analogous to the Network of Organic Chemistry, where each node is defined as a metabolite and the edges are enzyme-catalyzed reactions (Figure 4f) [195]. The maps are constructed using databases of metabolic reactions (see Table 1 for references). One of the most common is KEGG, which is metabolite centric and describes reactions generically by their EC number (a classification system organizing chemical reactions) [196].

Different networks can be used, depending on the engineering problem being addressed. For example, if the goal is to identify alternate routes through metabolism to increase flux through genetic engineering, then only the network associated with the chassis organism needs to be searched (*e.g.*, *E. coli* has 3755 metabolites) [197–203]. The goal could also be to define metabolic pathways from other organisms that could be moved into the target host [202,204–**206]. Then, all 18,505 metabolites and 11,146 reactions listed in KEGG irrespective of the source organism can be used to build a 'Universal Reaction Network' [207–*210]. When the target molecule is defined by the user, a search can then be performed from it to metabolites associated with the host organism [**45,211]. These approaches limit the user to select metabolites made either by the chassis

organism or known in biology, which are often the targets of metabolic engineering projects.

Rule-based approaches enable the expansion of the metabolic networks beyond empirically characterized reactions. Reaction rules are derived following approaches that are similar to those described for chemical retrosynthesis and may be manually curated or built automatically. Rules typically specify the reaction center and match the chemical structure a defined number of bonds away from that atom, by tallying the changes in bonds, or by changes in chemical features (roughly corresponding to the third digit of an EC number) [**44,**45,179,195,208,209,212–225]. To make more accurate predictions, the rules can be reduced by additional enzyme data, structural similarity between substrates, and molecular properties [**45,207,**213,226]. The molecular diversity of metabolites is less than the universe of organic chemistry, and this is reflected in the number of rules required. For example, only 250 rules capture most of the reactions in the KEGG database [195,227,228] and algorithms use as few as 50 [229], far smaller than the 60,000 rules in Chematica [**6].

The 'ATLAS of Biochemistry' and 'Metabolic *in silico* Network Expansion' databases were built using the complete set of hypothetical metabolites calculated on the basis of reaction rules [**220,**230]. Even these networks, at up to 571,000 nodes, are significantly smaller than the 35 million chemicals in the Network of Organic Chemistry. The connectivity is also less, with an average of $\sim$3 reactions per node, as opposed to $>$200 [231]. Thus, there is less need to develop advanced search algorithms. Small networks and short path searches can be performed through enumeration [198,229,232]. For example, 107,272 paths were found to isobutanol in *E. coli* [229]. Larger networks require branching algorithms to enumerate the pathways. The pruning steps in the network search can be decided on the basis of metrics that include the conservation of atoms between the substrate and product [200,201,205,233–235]. The organization of metabolic networks is different as compared to chemistry networks, particularly with the presence of hub compounds (ATP, NADPH, and so on) [199,200,203,231,236]. This can complicate a search, and several approaches have been taken to efficiently prune paths, for example, conserving functional groups to avoid short circuiting the search and creating unusable paths [234–237].

Objective functions are used to prioritize pathways for experiments [**44,**45,187,191,193,208,209,212,**215,238]. Most algorithms consider the shortest path as being advantageous as it requires the fewest heterologous enzymes and potentially toxic intermediates [187,211,**213,232,239,240]. Some paths have higher ATP requirements, thus reducing yield or lead to redox imbalances that can slow growth [**215,232]. The thermodynamic feasibility of pathways can be calculated on the basis of the predicted ΔG of each reaction [195,*210,238,**241–*243]. Toxic intermediates can also be avoided using predicted or empirical toxicity data [208,232,244–246]. Flux balance analysis is used to calculate the yield achievable with each path [211,**213,**215,232,247]. GEM-Path is able to calculate the yield under different oxygenation and growth conditions [**45].

The production of 1,4-butanediol (BDO) in *E. coli* is the only example of a retrosynthesis algorithm being used to design a *de novo* pathway to a xenobiotic compound [**44]. Genomatica used their in-house SimPheny Biopathway Predictor and a set of 50 reaction rules to identify 10,000 pathways between 4 and 6 genes to produce BDO from *E. coli* metabolites. These pathways were ranked on the basis of the number of steps, thermodynamic feasibility, and predicted yield. This guided the design of the pathway shown in Figure 2b and led to a strain that entered industrial production. While this software is not publicly available, several alternative algorithms have been able to reproduce this result and even suggest potential for improvement (Table 1) [**45,**215].

Most software packages only provide a set of reactions, for example, as EC numbers, and do not identify specific enzymes. GEM-Path supplies the user with the first enzyme homolog in its database that matches each reaction [**45]. RetroPath 2.0 considers the number of available enzyme sequences during pathway selection and has a metric for enzyme promiscuity as a measure of likelihood that an enzyme will catalyze a particular reaction [**206,**213]. This type of approach could be adapted for enzyme mining, where algorithms guide the sets of enzymes to be screened at each step. Taken a step further, data from enzyme mining efforts could be used to map enzymes to specific reactions and substrates. This approach has been taken within the chemical space of xenobiotic bisindoles in order to identify combinations of 21 enzymes obtained via part mining and used to create a barcoded DNA assembly scheme that allows a user to rapidly build a pathway to one of 540 potential target molecules (Figure 4g) [10].

Combining chemical and biochemical routes

The integration of chemical and biochemical steps will expand the molecule space accessible with either suite of techniques alone. Molecules that are trivial to chemistry, such as tetrahydrofuran, are difficult to make with biology because of toxicity or a lack of enzymes that perform the required reaction [10]. Conversely, chemistry struggles with regioselectivity and making modifications to large and highly functionalized molecules. Retrosynthesis algorithms could be developed that

divide a complex synthesis between chemical and biological steps, including those requiring multiple living cells, resting cells, and/or purified enzymes.

Increasingly, enzymatic steps are being introduced into chemical processes at an industrial scale [248–252]. As an example, nitrile hydratase is used to make 600,000 tons per year of acrylamide [253–256]. Merck developed a process to the pharmaceutical sitagliptin using an enzyme obtained by directed evolution to perform a challenging transamination (Figure 5a) [257,258]. Enzymes can be combined with each other and even metal catalysts to perform 1-pot reactions [257,259–261]. Similarly, there are techniques that treat cells as 'bags of enzyme' by pelleting cultures and resuspending them as resting cells or taking the extra disruptive steps to create freeze-dried cells or cell-free lysate. These approaches are cheap ways to implement a single or few enzymatic steps, rather than building a molecule from a sugar feedstock through central metabolism. These formulations are effectively treated as catalysts and have been combined with chemical and enzymatic steps, including in 1-pot reactions (Figure 5b) [54,262,263].

There are many ways to divide up chemical and biological steps in a retrosynthetic design. Chemical steps are often put before or after cellular processing. This can be to deal with a final step that produces a toxic product (e.g., to make artemisinin) or to feed cells substrates with chemical modifications difficult to produce enzymatically (e.g., to make halogenated benzylisoquinoline alkaloids) [**23,264]. Mixing and matching these approaches can lead to complex process schemes. For example, a patented route to caprilactam (a nylon-6 precursor) is to take lysine purified from cells, chemically convert it to aminohex-2-enoic acid, enzymatically convert that to 6-ACA, with a final chemical conversion to caprilactam [265]. There are many considerations in designing such processes, including the ability for substrates to cross the cell membrane, toxicity, and the balance between purification costs and decreased effectiveness in combining steps. Multiple species of living cells can be cocultured that perform different stages of the reaction scheme, for example, culturing engineered *E. coli* with yeast to make oxygenated taxanes or with *P. putida* to make carvolactone (Figure 2b and c) [*53,55]. The catechin pathway takes this to the extreme, with each of the four modules being contained in its own bacterial strain and using a coculture technique for production of the final product (Figure 2a) [25]. Substrates can be fed to cells that facilitate post-fermentation processing. For example, THC and bicyclobutanes were diversified by feeding cells substrates that can perform click chemistry (Figure 5c) [39,**122]. Currently, all these processes are put together in an *ad hoc* manner using expert knowledge. Algorithms will have to be developed that explore the potential paths and balance the numerous constraints in order to full harness the potential [266].

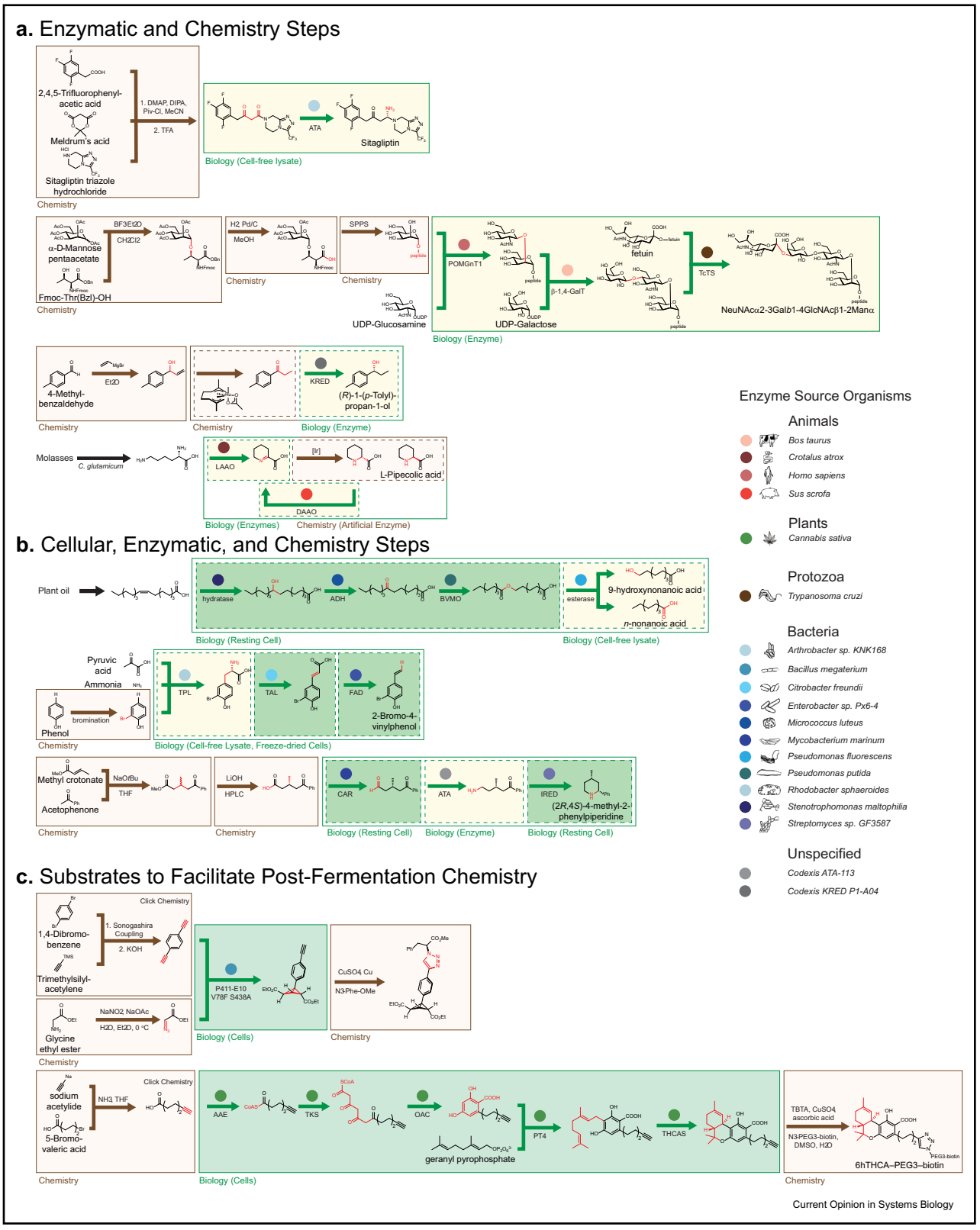
Advanced synthetic biology to facilitate retrosynthesis

This review has focused on the problem of retrosynthesis, that is, the identification of a set of enzymes that will convert a cellular metabolite into a xenobiotic product. Increasing the complexity of the molecules that can be produced will require additional tools. These include methods to increase the flux to the precursor metabolites and the control of when enzymes are expressed and where they are localized. This section focuses on anticipating those techniques required to build 40+ enzyme pathways to make a target molecule and not the additional work required to make the strain and process economically viable.

Strain development with the goal of enabling long retrosynthesis pathways may be subtly different compared to classical optimization for metabolic engineering projects. For example, both require increased flux to the precursor metabolite, but with differing goals. As a retrosynthetic pathway gets longer, each added enzyme is likely to decrease the amount of product made from the precursor. Thus, it is required to have the highest possible concentration of precursor metabolite available to the first enzyme. To enable a retrosynthesis project, this could come at the cost of decreasing yield from glucose or reducing the growth rate. This is a different objective than is typical in metabolic engineering projects, where optimizing yield and growth are critical [15,267–277]. For example, considering that many of the targets shown in Figure 2a are derived from tyrosine, one could imagine building strains of a platform chassis (*e.g.*, *Saccharomyces cerevisiae*) that produce the highest concentration of intracellular tyrosine possible without concern for its yield from glucose. This could be done for other commonly needed precursors [**45] to create a series of retrosynthesis-ready strains. Similarly, flux models and bioinformatics could be applied to identify and knockout native enzymes that are likely to modify chemicals derived from these precursors (as opposed to a specific pathway), thus reducing the likelihood of inadvertently siphoning intermediates away. For these efforts, selecting the right chassis is critical to avoid toxicity, express folded enzymes, or enhance precursor availability. Efforts in the field to create simplified genomes, develop fast-growing hosts, and tame diverse environmental organisms will provide more options in the future [15,278–286].

Central to organic chemistry is the concept of order of operations, where each reaction is performed as an independent step, and the product is purified before proceeding to the next reaction. While it is possible to combine steps into '1-pot' reactions, it is not possible to combine all those required for a complex retrosynthesis

Figure 5



Examples are shown of retrosynthetic efforts involving combinations of chemical and biological steps. The biological steps can involve enzymes in a cell-free context **(a)** or implemented by resting or living cells **(b)** [54,257,259–263]. Solid boxes indicate 1-pot reactions consisting of multiple simultaneous operations (dashed boxes). Enzyme colors indicate the source organisms of the enzymes (legend); 'unspecified' refers to enzymes that are part of kits provided by Codexis where the source information is not provided. **(c)** Two examples are shown where substrates are fed to cells that enable click chemistry to introduce further chemical modifications after isolating the product from fermentation [39,**122].

scheme (e.g., Figure 1b). Order of operations is unused in metabolic engineering: all enzymes are coexpressed all the time. In systems biology, there has been some characterization of the relationship between the timing of enzyme expression and metabolic cost for amino acid biosynthesis [287]. Order of operations is also encoded in megasynthases, but the ability to manipulate these systems remains elusive (much less harness for full retrosynthesis) [288–291]. Synthetic regulatory networks (gene circuits) have been used to switch enzyme expression on when the fermentation changes from the growth to production phase, reduce the accumulation of toxic intermediates, change the regulatory response to alternative feedstocks, and express enzymes under conditions when they are active [292–303]. A more advanced function would be to use them to time the expression of enzymes to build a molecule in steps, where the order of enzyme expression is required to build the target molecule. This could enable more complex pathway designs, for example, the biochemical equivalent of protection–deprotection as a means to avoid certain modifications by promiscuous enzymes.

Far from being bags of homogeneous biomolecules, cells offer means to spatially organize metabolic pathways. Naturally, localization is used to channel the molecule effectively between enzyme reaction sites, stop toxic intermediates from imparting their effect, or localize enzymes where the substrates are at their highest concentration [*304]. Megasynthases, for example polyketide synthases, are enzyme 'assembly lines' that progressively transform the product as it is passed between domains [288–290,305,306]. The modular nature of the enzyme sparked early efforts to implement retrosynthesis schemes by mixing-and-matching domains; while the understanding of this process has improved, this capability remains elusive [307–311]. Many copies of these enzymes can self-assemble into large (0.4 μm) organelle-like complexes localized at the cell membrane [312,313]. Various approaches have been developed to artificially scaffold enzymes by fusing them to binding domains that then assemble onto a protein, DNA, RNA, or lipid shell, that can then form larger structures [314–320]. These impart their effect by increasing the local concentration of intermediates and reducing the residence time of toxic intermediates [314,321].

Subcellular compartments, or organelles, can be used to separate enzymatic reaction steps. In eukaryotes, it is common practice to direct some enzymes in a pathway to organelles, such as mitochondria or the vacuole, to take advantage of higher concentrations of some metabolites or conditions that favor the activity of the participating enzymes (*e.g.*, salt or pH) [9,87,322]. Artificial organelles with a unique chemical signature have also been built in yeast by locally building up the synthesis of unused phospholipid species [323]. Membraneless organelles that form because of phase separation have been used to create an orthogonal translation system that can build proteins with noncanonical amino acids [324]. Prokaryotes are able to build phage-like microcompartments, into which enzymes are targeted, that have pores that selectively allow substrates in and sequester toxic intermediates [325–327]. Typically, the microcompartment genes occur together with metabolic genes as a cluster in the genome. The rules by which signal sequences can direct heterologous enzymes to the microcompartment and mutations can be made to the pore are being elucidated [325,328–335]. The prokaryotic microcompartment has been moved to yeast and incorporated into a biosynthetic pathway from tyrosine to norcoclaurine [336]. There are many ways that a target pathway could be sequestered via scaffolding proteins or divided up among organelles and microcompartments. Mathematical models have been developed in order to guide this decision-making [327,**337,338].

Conclusion

What is required to tame biology's ability to build molecules such that complex xenobiotic chemicals can be made by design? Even for simple targets, there have been only a tiny number of success stories (Figure 2b)—especially when compared to the enormous body of work from metabolic engineering efforts over the last few decades. We propose that this emerges from a tension between enzyme discovery favoring promiscuity but use in a pathway requiring specificity. Combining enzymes from different pathways, therefore, requires intensive efforts at each step to screen enzymes for those that perform the function required out-of-context with follow-up engineering efforts to deal with their promiscuity (either directly or indirectly). However, the field is at an inflection point of possibilities: enzyme mining, engineering, and evolution when combined with high-throughput metabolomics are going to lead to a deluge of specificity data. Algorithms for retrosynthetic design will integrate these data, to guide pathway construct to branch from a cellular metabolite to target molecule. New methods in artificial intelligence offer the possibility to integrate chemistry and biology into complex reaction schemes to provide access to complex molecules not achievable with either alone. Core capabilities in engineering cells need to be improved, including genome engineering, synthetic regulatory networks, and the construction of physical structures in cells. Achieving large retrosynthetic designs to build a xenobiotic molecule of a scale of a natural product is such a complex process that it will require integrated computer-aided design packages that combine retrosynthesis, metabolic flux analysis, protein engineering, and genetic circuit design automation.

Conflict of interest statement

Nothing declared.

Acknowledgements

This research was funded by the US Defense Advanced Research Projects Agency's Living Foundries program award HR0011-15-C-0084 and a research award from Novartis Institute for BioMedical Research (Cambridge, USA).

References

Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest
** of outstanding interest

1. Corey EJ: **General methods for the construction of complex molecules**. *Pure Appl Chem* 1967, **14**:19–37.

2. Aicher TD, Buszek KR, Fang FG, Forsyth CJ, Jung SH, Kishi Y, Matelich MC, Scola PM, Spero DM, Yoon SK: **Total synthesis of halichondrin B and norhalichondrin B**. *J Am Chem Soc* 1992, **114**:3162–3164.

3. Nicolaou KC, Sorensen EJ: *Classics in total synthesis: targets, strategies, methods*. Wiley-VCH; 1996.

4. Nicolaou KC, Snyder SA: *Classics in total synthesis II: more targets, strategies, methods*. Wiley-VCH; 2003.

5. Nicolaou KC, Chen JS: *Classics in total synthesis III: further targets, strategies, methods*. Wiley-VCH; 2011.

6. ** Szymkuć S, Gajewska EP, Klucznik T, Molga K, Dittwald P, Startek M, Bajczyk M, Grzybowski BA: **Computer-Assisted synthetic planning: the end of the beginning**. *Angew Chem Int Ed* 2016, **55**:5904–5937.

The paper provides an in-depth overview of the algorithm Chematica uses to navigate through the network of organic chemistry. It also discusses the human-curated rules in Chematica and their use for de novo retrosynthetic design.

7. Czar MJ, Anderson JC, Bader JS, Peccoud J: **Gene synthesis demystified**. *Trends Biotechnol* 2009, **27**:63–72.

8. Lee ME, DeLoache WC, Cervantes B, Dueber JE: **A highly characterized yeast toolkit for modular, multipart assembly**. *ACS Synth Biol* 2015, **4**:975–986.

9. Young EM, Zhao Z, Gielesen BEM, Wu L, Benjamin Gordon D, Roubos JA, Voigt CA: **Iterative algorithm-guided design of massive strain libraries, applied to itaconic acid production in yeast**. *Metab Eng* 2018, **48**:33–43.

10. Casini A, Chang F-Y, Eluere R, King AM, Young EM, Dudley QM, Karim A, Pratt K, Bristol C, Forget A, *et al.*: **A pressure test to make 10 molecules in 90 Days: external evaluation of methods to engineer biology**. *J Am Chem Soc* 2018, **140**:4302–4316.

11. Chao R, Yuan Y, Zhao H: **Building biological foundries for next-generation synthetic biology**. *Sci China Life Sci* 2015, **58**: 658–665.

12. Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison Iii CA, Smith HO: **Enzymatic assembly of DNA molecules up to several hundred kilobases**. *Nat Methods* 2009, **6**:343.

13. Engler C, Kandzia R, Marillonnet S: **A one pot, one step, precision cloning method with high throughput capability**. *PLoS One* 2008, **3**:e3647.

14. Canton B, Labno A, Endy D: **Refinement and standardization of synthetic biological parts and devices**. *Nat Biotechnol* 2008, **26**:787.

15. Brophy JAN, Triassi AJ, Adams BL, Renberg RL, Stratis-Cullum DN, Grossman AD, Voigt CA: **Engineered integrative and conjugative elements for efficient and inducible DNA transfer to undomesticated bacteria**. *Nature Microbiol* 2018, **3**: 1043–1053.

16. Bomgardner MM: **Ginkgo Bioworks and Zymergen scale up synthetic biology with robots**. In *Chemical & engineering news*; 2016:18–22.

17. Arnold FH: **Directed evolution: bringing new chemistry to life**. *Angew Chem Int Ed* 2018, **57**:4143–4148.

18. Zeymer C, Hilvert D: **Directed evolution of protein catalysts**. *Annu Rev Biochem* 2018, **87**:131–157.

19. Ro D-K, Paradise EM, Ouellet M, Fisher KJ, Newman KL, Ndungu JM, Ho KA, Eachus RA, Ham TS, Kirby J, *et al.*: **Production of the antimalarial drug precursor artemisinic acid in engineered yeast**. *Nature* 2006, **440**:940.

20. Paddon CJ, Westfall PJ, Pitera DJ, Benjamin K, Fisher K, McPhee D, Leavell MD, Tai A, Main A, Eng D, *et al.*: **High-level semi-synthetic production of the potent antimalarial artemisinin**. *Nature* 2013, **496**:528–532.

21. Edgar S, Li F-S, Qiao K, Weng J-K, Stephanopoulos G: **Engineering of taxadiene synthase for improved selectivity and yield of a key taxol biosynthetic intermediate**. *ACS Synth Biol* 2017, **6**:201–205.

22. Engels B, Dahm P, Jennewein S: **Metabolic engineering of taxadiene biosynthesis in yeast as a first step towards Taxol (Paclitaxel) production**. *Metab Eng* 2008, **10**:201–206.

23. ** Li Y, Li S, Thodey K, Trenchard I, Cravens A, Smolke CD: **Complete biosynthesis of noscapine and halogenated alkaloids in yeast**. *Proc Natl Acad Sci Unit States Am* 2018, **115**: E3922.

Yeast was engineered to produce noscapine and its analogs derived from halogenated tyrosine by expressing 25 heterologous enzymes from various organisms and 6 mutant or overexpressed endogenous yeast enzymes.

24. ** Galanie S, Thodey K, Trenchard IJ, Filsinger Interrante M, Smolke CD: **Complete biosynthesis of opioids in yeast**. *Science* 2015, **349**:1095.

Over 20 enzymes sourced from various organisms were expressed in yeast to reconstruct the opioid biosynthetic pathway, where enzyme mining, enzyme engineering, and optimization were employed to increase precursor supplies, improve bottlenecked steps, and facilitate cofactor recycling.

25. Jones JA, Vernacchio VR, Sinkoe AL, Collins SM, Ibrahim MHA, Lachance DM, Hahn J, Koffas MAG: **Experimental and computational optimization of an Escherichia coli co-culture for the efficient production of flavonoids**. *Metab Eng* 2016, **35**: 55–63.

26. Zhou H, Voigt CA, Vonk B, Roubos JA, Bovenberg RAL: **Algorithmic co-optimization of genetic constructs and growth conditions: application to 6-ACA, a potential nylon-6 precursor**. *Nucleic Acids Res* 2015, **43**:10560–10570.

27. Borodina I, Kildegaard KR, Jensen NB, Blicher TH, Maury J, Sherstyk S, Schneider K, Lamosa P, Herrgård MJ, Rosenstand I, *et al.*: **Establishing a synthetic pathway for high-level production of 3-hydroxypropionic acid in Saccharomyces cerevisiae via β-alanine**. *Metab Eng* 2015, **27**:57–64.

28. ** Curran KA, Leavitt JM, Karim AS, Alper HS: **Metabolic engineering of muconic acid production in Saccharomyces cerevisiae**. *Metab Eng* 2013, **15**:55–66.

In vitro enzyme characterization guided the selection and construction of non-natural muconic acid pathway in yeast whose titer was further improved by genetic engineering and balanced expression of heterologous enzymes.

29. Bang HB, Lee YH, Kim SC, Sung CK, Jeong KJ: **Metabolic engineering of Escherichia coli for the production of cinnamaldehyde**. *Microb Cell Factories* 2016, **15**:16.

30. Li J, Tian C, Xia Y, Mutanda I, Wang K, Wang Y: **Production of plant-specific flavones baicalein and scutellarein in an engineered E. coli from available phenylalanine and tyrosine**. *Metab Eng* 2019, **52**:124–133.

31. Kim M-J, Kim B-G, Ahn J-H: **Biosynthesis of bioactive O-methylated flavonoids in Escherichia coli**. *Appl Microbiol Biotechnol* 2013, **97**:7195–7204.

32. Wu J, Liu P, Fan Y, Bao H, Du G, Zhou J, Chen J: **Multivariate modular metabolic engineering of Escherichia coli to produce resveratrol from l-tyrosine**. *J Biotechnol* 2013, **167**: 404–411.

33. Wang S, Zhang S, Xiao A, Rasmussen M, Skidmore C, Zhan J: **Metabolic engineering of Escherichia coli for the biosynthesis of various phenylpropanoid derivatives**. *Metab Eng* 2015, **29**:153–159.

34. Slattery SS, Diamond A, Wang H, Therrien JA, Lant JT, Jazey T, Lee K, Klassen Z, Desgagné-Penix I, Karas BJ, *et al.*: **An expanded plasmid-based genetic toolbox enables Cas9 genome editing and stable maintenance of synthetic pathways in phaeodactylum tricornutum**. *ACS Synth Biol* 2018, **7**: 328–338.

35. Alonso-Gutierrez J, Chan R, Batth TS, Adams PD, Keasling JD, Petzold CJ, Lee TS: **Metabolic engineering of Escherichia coli for limonene and perillyl alcohol production**. *Metab Eng* 2013, **19**:33–41.

36. Zhang C, Chen X, Lindley ND, Too H-P: **A "plug-n-play" modular metabolic system for the production of apocarotenoids**. *Biotechnol Bioeng* 2018, **115**:174–183.

37. Brown S, Clastre M, Courdavault V, O'Connor SE: **De novo production of the plant-derived alkaloid strictosidine in yeast**. *Proc Natl Acad Sci Unit States Am* 2015, **112**:3205.

38. Nakagawa A, Minami H, Kim J-S, Koyanagi T, Katayama T, Sato F, Kumagai H: **A bacterial platform for fermentative production of plant alkaloids**. *Nat Commun* 2011, **2**:326.

39. Luo X, Reiter MA, d'Espaux L, Wong J, Denby CM, Lechner A, Zhang Y, Grzybowski AT, Harth S, Lin W, *et al.*: **Complete biosynthesis of cannabinoids and their unnatural analogues in yeast**. *Nature* 2019, **567**:123–126.

40. Gold ND, Fossati E, Hansen CC, DiFalco M, Douchin V, Martin VJJ: **A combinatorial approach to study cytochrome P450 enzymes for de novo production of steviol glucosides in Baker's yeast**. *ACS Synth Biol* 2018, **7**: 2918–2929.

41. Choi SY, Park SJ, Kim WJ, Yang JE, Lee H, Shin J, Lee SY: **One-step fermentative production of poly(lactate-co-glycolate) from carbohydrates in Escherichia coli**. *Nat Biotechnol* 2016, **34**:435.

42. Dhamankar H, Tarasova Y, Martin CH, Prather KLJ: **Engineering E. coli for the biosynthesis of 3-hydroxy-γ-butyrolactone (3HBL) and 3,4-dihydroxybutyric acid (3,4-DHBA) as value-added chemicals from glucose as a sole carbon source**. *Metab Eng* 2014, **25**:72–81.

43. Yang JE, Park SJ, Kim WJ, Kim HJ, Kim BJ, Lee H, Shin J, Lee SY: **One-step fermentative production of aromatic polyesters from glucose by metabolically engineered Escherichia coli strains**. *Nat Commun* 2018, **9**:79.

44. ** Yim H, Haselbeck R, Niu W, Pujol-Baxley C, Burgard A, Boldt J, Khandurina J, Trawick JD, Osterhout RE, Stephen R, *et al.*: **Metabolic engineering of Escherichia coli for direct production of 1,4-butanediol**. *Nat Chem Biol* 2011, **7**: 445–452.

De novo biosynthetic pathway for non-natural metabolite 1,4-butanediol was designed by an algorithm and realized experimentally in E. coli. The production was further increased by metabolic engineering guided by flux analysis.

45. ** Campodonico MA, Andrews BA, Asenjo JA, Palsson BO, Feist AM: **Generation of an atlas for commodity chemical production in Escherichia coli and a novel pathway prediction algorithm, GEM-Path**. *Metab Eng* 2014, **25**:140–158.

A target compound is selected and the algorithm finds all paths of heterologous enzymes that lead to the pool of metabolites in a defined host. The computational design of non-native pathways to 20 commodity chemicals in *E. coli* is demonstrated, where 245 reaction paths are dropped into genome-scale flux models and yield is calculated under different conditions.

46. Harris DM, Westerlaken I, Schipper D, van der Krogt ZA, Gombert AK, Sutherland J, Raamsdonk LM, van den Berg MA, Bovenberg RAL, Pronk JT, *et al.*: **Engineering of Penicillium chrysogenum for fermentative production of a novel carbamoylated cephem antibiotic precursor**. *Metab Eng* 2009, **11**: 125–137.

47. Sheppard MJ, Kunjapur AM, Wenck SJ, Prather KLJ: **Retrobiosynthetic screening of a modular pathway design achieves selective route for microbial synthesis of 4-methylpentanol**. *Nat Commun* 2014, **5**:5031.

48. Beekwilder J, van der Meer IM, Sibbesen O, Broekgaarden M, Qvist I, Mikkelsen JD, Hall RD: **Microbial production of natural raspberry ketone**. *Biotechnol J* 2007, **2**:1270–1279.

49. Minami H, Kim J-S, Ikezawa N, Takemura T, Katayama T, Kumagai H, Sato F: **Microbial production of plant benzylisoquinoline alkaloids**. *Proc Natl Acad Sci Unit States Am* 2008, **105**:7393.

50. Leonard E, Yan Y, Lim KH, Koffas MAG: **Investigation of two distinct flavone synthases for plant-specific flavone biosynthesis in Saccharomyces cerevisiae**. *Appl Environ Microbiol* 2005, **71**:8241.

51. Hwang EI, Kaneko M, Ohnishi Y, Horinouchi S: **Production of plant-specific flavanones by Escherichia coli containing an artificial gene cluster**. *Appl Environ Microbiol* 2003, **69**:2699.

52. Ajikumar PK, Xiao W-H, Tyo KEJ, Wang Y, Simeon F, Leonard E, Mucha O, Phon TH, Pfeifer B, Stephanopoulos G: **Isoprenoid pathway optimization for taxol precursor overproduction in Escherichia coli**. *Science* 2010, **330**:70.

53. * Zhou K, Qiao K, Edgar S, Stephanopoulos G: **Distributing a metabolic pathway among a microbial consortium enhances production of natural products**. *Nat Biotechnol* 2015, **33**: 377–383.

A theoretical framework is provided for comparing the benefits of putting a long pathway in one cell versus multiple cells.

54. Otte KB, Kittelberger J, Kirtz M, Nestl BM, Hauer B: **Whole-cell one-pot biosynthesis of azelaic acid**. *ChemCatChem* 2014, **6**: 1003–1009.

55. Oberleitner N, Ressmann AK, Bica K, Gärtner P, Fraaije MW, Bornscheuer UT, Rudroff F, Mihovilovic MD: **From waste to value – direct utilization of limonene from orange peel in a biocatalytic cascade reaction towards chiral carvolactone**. *Green Chem* 2017, **19**:367–371.

56. Wu S, Zhou Y, Wang T, Too H-P, Wang DIC, Li Z: **Highly regio- and enantioselective multiple oxy- and amino-functionalizations of alkenes by modular cascade biocatalysis**. *Nat Commun* 2016, **7**. 11917-11917.

57. * Thodey K, Galanie S, Smolke CD: **A microbial biomanufacturing platform for natural and semisynthetic opioids**. *Nat Chem Biol* 2014, **10**:837.

Yeast was engineered to convert thebaine to morphine and its derivatives by expression of plant and bacterial enzymes, which was further improved by supplementation of precursor, enzyme balance, and spatial engineering.

58. Bruce NC, Wilmot CJ, Jordan KN, Stephens LDG, Lowe CR: **Microbial degradation of the morphine alkaloids. Purification and characterization of morphine dehydrogenase from Pseudomonas putida M10**. *Biochem J* 1991, **274**:875.

59. Hawkins KM, Smolke CD: **Production of benzylisoquinoline alkaloids in Saccharomyces cerevisiae**. *Nat Chem Biol* 2008, **4**:564.

60. Dekishima Y, Lan EI, Shen CR, Cho KM, Liao JC: **Extending carbon chain length of 1-butanol pathway for 1-hexanol synthesis from glucose by engineered Escherichia coli**. *J Am Chem Soc* 2011, **133**:11399–11401.

61. Zirpel B, Stehle F, Kayser O: **Production of Δ9-tetrahydrocannabinolic acid from cannabigerolic acid by whole cells of Pichia (Komagataella) pastoris expressing Δ9-tetrahydrocannabinolic acid synthase from Cannabis sativa I**. *Biotechnol Lett* 2015, **37**:1869–1875.

62. Zirpel B, Degenhardt F, Martin C, Kayser O, Stehle F: **Engineering yeasts as platform organisms for cannabinoid biosynthesis**. *J Biotechnol* 2017, **259**:204–212.

63. van Bakel H, Stout JM, Cote AG, Tallon CM, Sharpe AG, Hughes TR, Page JE: **The draft genome and transcriptome of Cannabis sativa**. *Genome Biol* 2011, **12**:R102.

64. Carvalho Â, Hansen EH, Carlsen S, Stehle F, Kayser O: **Designing microorganisms for heterologous biosynthesis of cannabinoids**. *FEMS Yeast Res* 2017, **17**.

65. Biz A, Proulx S, Xu Z, Siddartha K, Mulet Indrayanti A, Mahadevan R: **Systems biology based metabolic engineering for non-natural chemicals**. *Biotechnol Adv* 2019.

66. Liu H, Lu T: **Autonomous production of 1,4-butanediol via a de novo biosynthesis pathway in engineered Escherichia coli**. *Metab Eng* 2015, **29**:135–141.

67. Wang J, Jain R, Shen X, Sun X, Cheng M, Liao JC, Yuan Q, Yan Y: **Rational engineering of diol dehydratase enables 1,4-butanediol biosynthesis from xylose**. *Metab Eng* 2017, **40**: 148–156.

68. Tseng H-C, Prather KLJ: **Controlled biosynthesis of odd-chain fuels and chemicals via engineered modular metabolic pathways**. *Proc Natl Acad Sci Unit States Am* 2012, **109**:17925.

69. Kang S-Y, Choi O, Lee JK, Ahn J-O, Ahn JS, Hwang BY, Hong Y-S: **Artificial de novo biosynthesis of hydroxystyrene derivatives in a tyrosine overproducing Escherichia coli strain**. *Microb Cell Factories* 2015, **14**:78.

70. Furubayashi M, Ikezumi M, Takaichi S, Maoka T, Hemmi H, Ogawa T, Saito K, Tobias AV, Umeno D: **A highly selective biosynthetic pathway to non-natural C50 carotenoids assembled from moderately selective enzymes**. *Nat Commun* 2015, **6**:7534.

71. McKenna R, Pugh S, Thompson B, Nielsen DR: **Microbial production of the aromatic building-blocks (S)-styrene oxide and (R)-1,2-phenylethanediol from renewable resources**. *Biotechnol J* 2013, **8**:1465–1475.

72. * Turk SCHJ, Kloosterman WP, Ninaber DK, Kolen KPAM, Knutova J, Suir E, Schürmann M, Raemakers-Franken PC, Müller M, de Wildeman SMA, *et al.*: **Metabolic engineering toward sustainable production of nylon-6**. *ACS Synth Biol* 2016, **5**:65–73.

The 6-ACA pathway required the identification of five novel enzymatic activities, which was done by using DNA synthesis to construct sets of enzymes that act on similar substrates (15 aldehyde dehydrogenases, 6 and 4 aminotransferases for two steps, 4 and 3 decarboxylases for two steps).

73. Smit BA, van Hylckama Vlieg JET, Engels WJM, Meijer L, Wouters JTM, Smit G: **Identification, cloning, and characterization of a Lactococcus lactis branched-chain α-keto acid decarboxylase involved in flavor formation**. *Appl Environ Microbiol* 2005, **71**:303.

74. Zheng L, White RH, Dean DR: **Purification of the Azotobacter vinelandii nifV-encoded homocitrate synthase**. *J Bacteriol* 1997, **179**:5963.

75. Shin JS, Yun H, Jang JW, Park I, Kim BG: **Purification, characterization, and molecular cloning of a novel amine:pyruvate transaminase from Vibrio fluvialis JS17**. *Appl Microbiol Biotechnol* 2003, **61**:463–471.

76. Drevland RM, Jia Y, Palmer DRJ, Graham DE: **Methanogen homoaconitase catalyzes both hydrolyase reactions in coenzyme B biosynthesis**. *J Biol Chem* 2008, **283**:28888–28896.

77. Landenmark HKE, Forgan DH, Cockell CS: **An estimate of the total DNA in the biosphere**. *PLoS Biol* 2015, **13**:e1002168.

78. Agarwala R, Barrett T, Beck J, Benson DA, Bollin C, Bolton E, Bourexis D, Brister JR, Bryant SH, Canese K, *et al.*: **Database resources of the national center for biotechnology information**. *Nucleic Acids Res* 2018, **46**:D8–D13.

79. Tian W, Arakaki AK, Skolnick J: **EFICAz: a comprehensive approach for accurate genome-scale enzyme function inference**. *Nucleic Acids Res* 2004, **32**:6226–6239.

80. Newton MS, Arcus VL, Gerth ML, Patrick WM: **Enzyme evolution: innovation is easy, optimization is complicated**. *Curr Opin Struct Biol* 2018, **48**:110–116.

81. Lipman DJ, Benson DA, Karsch-Mizrachi I, Ostell J, Clark K, Cavanaugh M, Sayers EW: **GenBank**. *Nucleic Acid Res* 2016, **45**:D37–D42.

82. * **NCBI reference sequence (RefSeq) database distribution release notes**. *Release* 2019, **93**.

This estimate is based on the 135,670,032 proteins in the NCBI database in 2019 and the estimation that 30% are enzymes from Tian et al. EFICAz: a comprehensive approach for accurate genome-scale enzyme function inference. *Nucleic Acids Research* 2004, 32:6226–6239.

83. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ, Pennsylvania T, Park U: **Basic local alignment search tool 2Department of computer science**. *J Mol Biol* 1990, **215**: 403–410.

84. Finn RD, Clements J, Eddy SR: **HMMER web server: interactive sequence similarity searching**. *Nucleic Acids Res* 2011, **39**: W29–W37.

85. Bateman A, Martin Maria J, O'Donovan C, Magrane M, Alpi E, Antunes R, Bely B, Bingley M, Bonilla C, Britto R, *et al.*: **UniProt: the universal protein knowledgebase**. *Nucleic Acids Res* 2017, **45**:D158–D169.

86. Atsumi S, Hanai T, Liao JC: **Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels**. *Nature* 2008, **451**:86–89.

87. Bayer TS, Widmaier DM, Temme K, Mirsky EA, Santi DV, Voigt CA: **Synthesis of methyl halides from biomass using engineered microbes**. *J Am Chem Soc* 2009, **131**:6508–6515.

88. Shen CR, Lan EI, Dekishima Y, Baez A, Cho KM, Liao JC: **Driving forces enable high-titer anaerobic 1-butanol synthesis in Escherichia coli**. *Appl Environ Microbiol* 2011, **77**: 2905–2915.

89. Heins RA, Cheng X, Nath S, Deng K, Bowen BP, Chivian DC, Datta S, Friedland GD, D'Haeseleer P, Wu D, *et al.*: **Phylogenomically guided identification of industrially relevant GH1 β-glucosidases through DNA synthesis and nanostructure-initiator mass spectrometry**. *ACS Chem Biol* 2014, **9**: 2082–2091.

90. Borodina I, Kildegaard KR, Jensen NB, Blicher TH, Maury J, Sherstyk S, Schneider K, Lamosa P, Herrgård MJ, Rosenstand I, *et al.*: **Establishing a synthetic pathway for high-level production of 3-hydroxypropionic acid in Saccharomyces cerevisiae via β-alanine**. *Metab Eng* 2014, **27**:57–64.

91. Ding Mz, Yan Hf, Li L-f, Zhai F, Shang L-q, Yin Z, Yuan Y-j: **Biosynthesis of taxadiene in Saccharomyces cerevisiae : selection of geranylgeranyl diphosphate synthase directed by a computer-aided docking strategy**. *PLoS One* 2014, **9**. e109348-e109348.

92. Jendresen CB, Stahlhut SG, Li M, Gaspar P, Siedler S, Förster J, Maury J, Borodina I, Nielsen AT: **Highly active and specific tyrosine ammonia-lyases from diverse origins enable enhanced production of aromatic compounds in bacteria and Saccharomyces cerevisiae**. *Appl Environ Microbiol* 2015, **81**: 4458–4476.

93. Eichenberger M, Lehka BJ, Folly C, Fischer D, Martens S, Simón E, Naesby M: **Metabolic engineering of Saccharomyces cerevisiae for de novo production of dihydrochalcones with known antioxidant, antidiabetic, and sweet tasting properties**. *Metab Eng* 2017, **39**:80–89.

94. Freund GS, O'Brien TE, Vinson L, Carlin DA, Yao A, Mak WS, Tagkopoulos I, Facciotti MT, Tantillo DJ, Siegel JB: **Elucidating substrate promiscuity within the FabI enzyme family**. *ACS Chem Biol* 2017, **12**:2465–2473.

95. Guggenheim KG, Crawford LM, Paradisi F, Wang SC, Siegel JB: **β-Glucosidase discovery and design for the degradation of oleuropein**. *ACS Omega* 2018, **3**:15754–15762.

96. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M: **Data, information, knowledge and principle: back to metabolism in KEGG**. *Nucleic Acids Res* 2014, **42**: 199–205.

97. Caspi R, Billington R, Ferrer L, Foerster H, Fulcher CA, Keseler IM, Kothari A, Krummenacker M, Latendresse M, Mueller LA, *et al.*: **The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases**. *Nucleic Acids Res* 2016, **44**:D471–D480.

98. Altman T, Travers M, Kothari A, Caspi R, Karp PD: **A systematic comparison of the MetaCyc and KEGG pathway databases**. *BMC Bioinf* 2013, **14**:112.

99. Chang A, Schomburg I, Jeske L, Placzek S, Schomburg D: **BRENDA in 2019: a European ELIXIR core data resource**. *Nucleic Acids Res* 2018, **47**:D542–D549.

100. Medema MH, Kottmann R, Yilmaz P, Cummings M, Biggins JB, Blin K, De Bruijn I, Chooi YH, Claesen J, Coates RC, *et al.*: **Minimum information about a biosynthetic gene cluster**. *Nat Chem Biol* 2015, **11**:625–631.

101. Blin K, Weber T, Lee SY, Medema MH, Pascal Andreu V, de los Santos EL C, Del Carratore F: **The antiSMASH database version 2: a comprehensive resource on secondary metabolite biosynthetic gene clusters**. *Nucleic Acids Res* 2018, **47**:D625–D630.

102. ** Blin K, Wolf T, Chevrette MG, Lu X, Schwalen CJ, Kautsar SA, Suarez Duran HG, de los Santos Emmanuel LC, Kim HU, Nave M, *et al.*: **antiSMASH 4.0—improvements in chemistry prediction and gene cluster boundary identification**. *Nucleic Acids Res* 2017, **45**:W36–W41.
AntiSMASH analyzes proteins encoded in the DNA sequence using profile Hidden Markov Models and makes bacterial / fungal biosynthetic gene cluster prediction based on a set of man-curated rules.

103. Gerlt JA, Bouvier JT, Davidson DB, Imker HJ, Sadkhin B, Slater DR, Whalen KL: **Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): a web tool for generating protein sequence similarity networks**. *Biochim Biophys Acta Protein Proteonomics* 2015, **1854**:1019–1037.

104. ** Mak WS, Tran S, Marcheschi R, Bertolani S, Thompson J, Baker D, Liao JC, Siegel JB: **Integrative genomic mining for enzyme function to enable engineering of a non-natural biosynthetic pathway**. *Nat Commun* 2015, **6**:1–9.
In searching for ketoisovalerate decarboxylase homologs that prefer C8 substrates ligand-enzyme models using Rosetta Design effectively prioritized 10 out of 239 candidates.

105. * Vanacek P, Sebestova E, Babkova P, Bidmanova S, Daniel L, Dvorak P, Stepankova V, Chaloupkova R, Brezovsky J, Prokop Z, *et al.*: **Exploration of enzyme diversity by integrating bioinformatics with expression analysis and biochemical characterization**. *ACS Catal* 2018, **8**:2402–2412.
An enzyme mining effort where homology models guided the calculation of ligand binding pocket volume. This was used to narrow down 658 dehalogenases to a set of 20 that are screened for their substrate profile using robotic automation.

106. Zhao S, Kumar R, Sakai A, Vetting MW, Wood BM, Brown S, Bonanno JB, Hillerich BS, Seidel RD, Babbitt PC, *et al.*: **Discovery of new enzymes and metabolic pathways by using structure and genome context**. *Nature* 2013, **502**:698–702.

107. Kosuri S, Church GM: **Large-scale de novo DNA synthesis: technologies and applications**. *Nat Methods* 2014, **11**:499–507.

108. Carlson R: *Time for new DNA synthesis and sequencing cost curves*. 2014. https://synbiobeta.com/time-new-dna-synthesis-sequencing-cost-curves-rob-carlson/.

109. ** Obexer R, Godina A, Garrabou X, Mittl PRE, Baker D, Griffiths AD, Hilvert D: **Emergence of a catalytic tetrad during evolution of a highly active artificial aldolase**. *Nat Chem* 2016, **9**:50.
An ultrahigh-throughput droplet-based microfluidic screening was applied to improve the activity of a computationally designed aldolase.

110. Dörr M, Fibinger MPC, Last D, Schmidt S, Santos-Aberturas J, Böttcher D, Hummel A, Vickers C, Voss M, Bornscheuer UT: **Fully automatized high-throughput enzyme library screening using a robotic platform**. *Biotechnol Bioeng* 2016, **113**:1421–1432.

111. Chen B, Lim S, Kannan A, Alford SC, Sunden F, Herschlag D, Dimov IK, Baer TM, Cochran JR: **High-throughput analysis and protein engineering using microcapillary arrays**. *Nat Chem Biol* 2015, **12**:76.

112. Abécassis V, Urban P, Aggerbeck L, Truan G, Pompon D: **Exploration of natural and artificial sequence spaces: towards a functional remodeling of membrane-bound cytochrome P450**. *Biocatal Biotransform* 2003, **21**:55–66.

113. Mair P, Gielen F, Hollfelder F: **Exploring sequence space in search of functional enzymes using microfluidic droplets**. *Curr Opin Chem Biol* 2017, **37**:137–144.

114. Ma F, Chung MT, Yao Y, Nidetz R, Lee LM, Liu AP, Feng Y, Kurabayashi K, Yang G-Y: **Efficient molecular evolution to generate enantioselective enzymes using a dual-channel microfluidic droplet screening platform**. *Nat Commun* 2018, **9**:1030.

115. Bunzel HA, Garrabou X, Pott M, Hilvert D: **Speeding up enzyme discovery and engineering with ultrahigh-throughput methods**. *Curr Opin Struct Biol* 2018, **48**:149–156.

116. ** Gielen F, Hours R, Emond S, Fischlechner M, Schell U, Hollfelder F: **Ultrahigh-throughput–directed enzyme evolution by absorbance-activated droplet sorting (AADS)**. *Proc Natl Acad Sci Unit States Am* 2016, **113**:E7383–E7389.
A NAD^+ dependent dye assay enables the assaying of >1,000,000 mutants per hour using a sorter.

117. Wahler D, Badalassi F, Crotti P, Reymond J-L: **Enzyme fingerprints of activity, and stereo- and enantioselectivity from fluorogenic and chromogenic substrate arrays**. *Chem Eur J* 2002, **8**:3211–3228.

118. Grognux J, Reymond J-L: **Classifying enzymes from selectivity fingerprints**. *Chembiochem* 2004, **5**:826–831.

119. Zhu Q, Uttamchandani M, Li D, Lesaicherre ML, Yao SQ: **Enzymatic profiling system in a small-molecule microarray**. *Org Lett* 2003, **5**:1257–1260.

120. ** Sévin DC, Fuhrer T, Zamboni N, Sauer U: **Nontargeted in vitro metabolomics for high-throughput identification of novel enzymes in Escherichia coli**. *Nat Methods* 2016, **14**:187.
Metabolome extract of *E. coli* was treated with cell lysate from *E. coli* expressing the target enzyme or its purified form. Non-targeted mass spectrometry was used to detect the change in metabolomes and assign function to 214 out of 1275 uncharacaterized *E. coli* enzymes.

121. Chen K, Huang X, Kan SBJ, Zhang RK, Arnold FH: **Enzymatic construction of highly strained carbocycles**. *Science* 2018, **360**:71–75.

122. ** Kan SBJ, Lewis RD, Chen K, Arnold FH: **Directed evolution of cytochrome c for carbon-silicon bond formation: bringing silicon to life**. *Science* 2016, **354**:1048–1051.
Cytochrome c from *Rhodothermus marinus* was enigeered to catalyze highly selective carbene insertion into silicon–hydrogen bond on various organosilicon compounds both *in vivo* and *in vitro*.

123. Copley SD: **Shining a light on enzyme promiscuity**. *Curr Opin Struct Biol* 2017, **47**:167–175.

124. Tawfik OKaDS: **Enzyme promiscuity: a mechanistic and evolutionary perspective**. *Annu Rev Biochem* 2010, **79**:471–505.

125. May O, Voigt CA, Arnold FH: **Enzyme engineering by directed evolution**. In *Enzyme catalysis in organic synthesis*. Edited by Drauz K, Waldmann H, Wiley-VCH; 2002:95–138.

126. ** Huang H, Pandya C, Liu C, Al-Obaidi NF, Wang M, Zheng L, Toews Keating S, Aono M, Love JD, Evans B, *et al.*: **Panoramic view of a superfamily of phosphatases through substrate profiling**. *Proc Natl Acad Sci Unit States Am* 2015, **112**:E1974–E1983.
Representative 217 phosphatases from diverse prokaryote species were expressed and their activity against a set of 167 substrates tested. The results revealed the different substrate specificity among these enzymes, which could be used to infer enzyme functions for unchacaterized phosphates.

127. Fasan R, Meharenna YT, Snow CD, Poulos TL, Arnold FH: **Evolutionary history of a specialized P450 propane monooxygenase**. *J Mol Biol* 2008, **383**:1069–1080.

128. Junker S, Roldan R, Joosten H-J, Clapés P, Fessner W-D: **Complete switch of reaction specificity of an aldolase by directed evolution in vitro: synthesis of generic aliphatic aldol products**. *Angew Chem Int Ed* 2018, **57**:10153–10157.

129. Colin P-Y, Kintses B, Gielen F, Miton CM, Fischer G, Mohamed MF, Hyvönen M, Morgavi DP, Janssen DB, Hollfelder F: **Ultrahigh-throughput discovery of promiscuous**

enzymes by picodroplet functional metagenomics. *Nat Commun* 2015, **6**:10008.

130. Khanal A, Kershner JP, Yu McLoughlin S, Copley SD: **Differential effects of a mutation on the normal and promiscuous activities of orthologs: implications for natural and directed evolution**. *Mol Biol Evol* 2014, **32**:100–108.

131. Rahimi M, van der Meer J-Y, Geertsema EM, Poddar H, Baas B-J, Poelarends GJ: **Mutations closer to the active site improve the promiscuous aldolase activity of 4-oxalocrotonate tautomerase more effectively than distant mutations**. *Chembiochem* 2016, **17**:1225–1228.

132. Marsden AFA, Wilkinson B, Cortés J, Dunster NJ, Staunton J, Leadlay PF: **Engineering broader specificity into an antibiotic-producing polyketide synthase**. *Science* 1998, **279**:199.

133. Jestin J-L, Vichier-Guerre S: **How to broaden enzyme substrate specificity: strategies, implications and applications**. *Res Microbiol* 2005, **156**:961–966.

134. Murphy AC, Hong H, Vance S, Broadhurst RW, Leadlay PF: **Broadening substrate specificity of a chain-extending ketosynthase through a single active-site mutation**. *Chem Commun* 2016, **52**:8373–8376.

135. Gupta RD: **Recent advances in enzyme promiscuity**. *Sustain Chem Proc* 2016, **4**:2.

136. Bornscheuer UT, Kazlauskas RJ: **Catalytic promiscuity in biocatalysis: using old enzymes to form new bonds and follow new pathways**. *Angew Chem Int Ed* 2004, **43**:6032–6040.

137. Schmidt DMZ, Mundorff EC, Dojka M, Bermudez E, Ness JE, Govindarajan S, Babbitt PC, Minshull J, Gerlt JA: **Evolutionary potential of (β/α)8-Barrels: functional promiscuity produced by single substitutions in the enolase superfamily**. *Biochemistry* 2003, **42**:8387–8393.

138. Chen R, Gao B, Liu X, Ruan F, Zhang Y, Lou J, Feng K, Wunsch C, Li S-M, Dai J, *et al.*: **Molecular insights into the enzyme promiscuity of an aromatic prenyltransferase**. *Nat Chem Biol* 2016, **13**:226.

139. Wijma HJ, Floor RJ, Bjelic S, Marrink SJ, Baker D, Janssen DB: **Enantioselective enzymes by computational design and in silico screening**. *Angew Chem Int Ed* 2015, **54**:3726–3730.

140. Ekroos M, Sjögren T: **Structural basis for ligand promiscuity in cytochrome P450 3A4**. *Proc Natl Acad Sci Unit States Am* 2006, **103**:13682–13687.

141. Hughes TB, Dang NL, Miller GP, Swamidass SJ: **Modeling reactivity to biological macromolecules with a deep multitask network**. *ACS Cent Sci* 2016, **2**:529–537.

142. Zaretzki J, Matlock M, Swamidass SJ: **XenoSite: accurately predicting CYP-mediated sites of metabolism with neural networks**. *J Chem Inf Model* 2013, **53**:3373–3383.

143. Tian S, Djoumbou-Feunang Y, Greiner R, Wishart DS: **CypReact: a software tool for in silico reactant prediction for human cytochrome P450 enzymes**. *J Chem Inf Model* 2018, **58**:1282–1291.

144. Bezhentsev VM, Tarasova OA, Dmitriev AV, Rudik AV, Lagunin AA, Filimonov DA, Poroikov VV: **Computer-aided prediction of xenobiotic metabolism in the human body**. *Russian Chem Rev* 2016, **85**:854.

145. Branneby C, Carlqvist P, Magnusson A, Hult K, Brinck T, Berglund P: **Carbon–Carbon bonds by hydrolytic enzymes**. *J Am Chem Soc* 2003, **125**:874–875.

146. Packer MS, Liu DR: **Methods for the directed evolution of proteins**. *Nat Rev Genet* 2015, **16**:379–394.

147. Currin A, Swainston N, Day PJ, Kell DB: **Synthetic biology for the directed evolution of protein biocatalysts: navigating sequence space intelligently**. *Chem Soc Rev* 2015, **44**:1172–1239.

148. Denard CA, Ren H, Zhao H: **Improving and repurposing biocatalysts via directed evolution**. *Curr Opin Chem Biol* 2015, **25**:55–64.

149. Hammer SC, Knight AM, Arnold FH: **Design and evolution of enzymes for non-natural chemistry**. *Curr Opin Green Sustain Chem* 2017, **7**:23–30.

150. ** Coelho PS, Brustad EM, Kannan A, Arnold FH: **Olefin cyclopropanation via carbene transfer catalyzed by engineered cytochrome P450 enzymes**. *Science* 2013, **339**:307–310.
$P450_{BM3}$ was engineered to catalyze diastereo- and stereoselective cyclopropanation on variou.

151. ** Jennifer Kan SB, Huang X, Gumulya Y, Chen K, Arnold FH: **Genetically programmed chiral organoborane synthesis**. *Nature* 2017, **552**:132–136.
Cytochrome c from *Rhodothermus marinus* was enigeered to catalyze highly selective carbene insertion into boron–hydrogen bond on various *N*-heterocyclic carbene borane compounds both *in vivo* and *in vitro*.

152. Röthlisberger D, Khersonsky O, Wollacott AM, Jiang L, DeChancie J, Betker J, Gallaher JL, Althoff EA, Zanghellini A, Dym O, *et al.*: **Kemp elimination catalysts by computational enzyme design**. *Nature* 2008, **453**:190–195.

153. Jiang L, Althoff EA, Clemente FR, Doyle L, Rothlisberger D, Zanghellini A, Gallaher JL, Betker JL, Tanaka F, Barbas CF, *et al.*: **De novo computational design of retro-aldol enzymes**. *Science* 2008, **319**:1387–1391.

154. Siegel JB, Zanghellini A, Lovick HM, Kiss G, Lambert AR, St Clair JL, Gallaher JL, Hilvert D, Gelb MH, Stoddard BL, *et al.*: **Computational design of an enzyme catalyst for a stereoselective bimolecular diels-alder reaction**. *Science* 2010, **329**:309–313.

155. Zanghellini A: **De novo computational enzyme design**. *Curr Opin Biotechnol* 2014, **29**:132–138.

156. Kim J, Kershner JP, Novikov Y, Shoemaker RK, Copley SD: **Three serendipitous pathways in E. coli can bypass a block in pyridoxal-5′-phosphate synthesis**. *Mol Syst Biol* 2010, **6**:436.

157. Faulon J-L, Carbonell P: **Molecular signatures-based prediction of enzyme promiscuity**. *Bioinformatics* 2010, **26**:2012–2019.

158. Corey EJ, Wipke WT, Cramer RD, Howe WJ: **Computer-assisted synthetic analysis. Facile man-machine communication of chemical structure by interactive computer graphics**. *J Am Chem Soc* 1972, **94**:421–430.

159. Corey EJ, Wipke WT: **Computer-Assisted design of complex organic syntheses**. *Science* 1969, **166**:178–192.

160. ** Klucznik T, Mikulak-Klucznik B, McCormack MP, Lima H, Szymkuć S, Bhowmick M, Molga K, Zhou Y, Rickershauser L, Gajewska EP, *et al.*: **Efficient syntheses of diverse, medicinally relevant targets planned by computer and executed in the laboratory**. *Chem* 2018, **4**:522–532.
Chematica was used to design a new synthetic routes for eight valuable chemicals. These routes are more efficient than those reported in literature and offer higher yield at reduced costs.

161. ** Segler MHS, Preuss M, Waller MP: **Planning chemical syntheses with deep neural networks and symbolic AI**. *Nature* 2018, **555**:604–610.
Based on rules extracted from Reaxys, a retrosynthesis algorithm uses three neural networks that predict the best reactions to make a product, limit reactions to those that are feasible, and estimate the position value. The tree is searched using a Monte Carlo algorithm.

162. Schwaller P, Gaudin T, Lányi D, Bekas C, Laino T: **"Found in Translation": predicting outcomes of complex organic chemistry reactions using neural sequence-to-sequence models**. *Chem Sci* 2018, **9**:6091–6098.

163. Baskin II , Madzhidov TI, Antipin IS, Varnek AA: **Artificial intelligence in synthetic chemistry: achievements and prospects**. *Russian Chem Rev* 2017, **86**:1127–1156.

164. Coley CW, Barzilay R, Jaakkola TS, Green WH, Jensen KF: **Prediction of organic reaction outcomes using machine learning**. *ACS Cent Sci* 2017, **3**:434–443.

165. Kayala MA, Azencott C-A, Chen JH, Baldi P: **Learning to predict chemical reactions**. *J Chem Inf Model* 2011, **51**:2209–2222.

166. * Liu B, Ramsundar B, Kawthekar P, Shi J, Gomes J, Luu Nguyen Q, Ho S, Sloane J, Wender P, Pande V: **Retrosynthetic reaction prediction using neural sequence-to-sequence models**. *ACS Cent Sci* 2017, **3**:1103–1113.
Single retrosynthetic steps are predicted using a recurrent neural network (RNN) that is trained to predict precursor molecules for 10 reaction types.

167. Kayala MA, Baldi P: **Reaction Predictor: prediction of complex chemical reactions at the mechanistic level using machine learning**. *J Chem Inf Model* 2012, **52**:2526–2540.

168. Coley CW, Green WH, Jensen KF: **Machine learning in computer-aided synthesis planning**. *Acc Chem Res* 2018, **51**: 1281–1289.

169. Peiretti F, Brunel JM: **Artificial intelligence: the future for organic chemistry?** *ACS Omega* 2018, **3**:13263–13266.

170. Carrera GVSM, Gupta S, Aires-de-Sousa J: **Machine learning of chemical reactivity from databases of organic reactions**. *J Comput Aided Mol Des* 2009, **23**:419–429.

171. Fialkowski M, Bishop KJM, Chubukov VA, Campbell CJ, Grzybowski BA: **Architecture and evolution of organic chemistry**. *Angew Chem Int Ed* 2005, **44**:7263–7269.

172. Bishop KJM, Klajn R, Grzybowski BA: **The core and most useful molecules in organic chemistry**. *Angew Chem Int Ed* 2006, **45**: 5348–5354.

173. Reaxys.https://www.reaxys.com/.

174. Jacob P-M, Lapkin A: **Statistics of the network of organic chemistry**. *React Chem Eng* 2018, **3**:102–118.

175. Feng F, Lai L, Pei J: **Computational chemical synthesis analysis and pathway design**. *Front Chem* 2018, **6**. 199-199.

176. Kowalik M, Gothard CM, Drews AM, Gothard NA, Weckiewicz A, Fuller PE, Grzybowski BA, Bishop KJM: **Parallel optimization of synthetic pathways within the network of organic chemistry**. *Angew Chem Int Ed* 2012, **51**:7928–7932.

177. * Grzybowski BA, Bishop KJM, Kowalczyk B, Wilmer CE: **The 'wired' universe of organic chemistry**. *Nat Chem* 2009, **1**:31–36.
The Network of Organic Chemistry is searched to identify chemicals of broad utility and to identify "islands" to which there is no synthetic route (typically natural products).

178. ** Gothard CM, Soh S, Gothard NA, Kowalczyk B, Wei Y, Baytekin B, Grzybowski BA: **Rewiring chemistry: algorithmic discovery and experimental validation of one-pot reactions in the network of organic chemistry**. *Angew Chem Int Ed* 2012, **51**:7922–7927.
The Network of Chemistry is searched to identify combinations of reactions that can be carried out under the same conditions (1 pot) and this is validated by synthesizing a pharmaceutical requiring four sequential reactions (cyclization, chlorination, alkynylation, and arylation) in one setup.

179. Law J, Zsoldos Z, Simon A, Reid D, Liu Y, Khew SY, Johnson AP, Major S, Wade RA, Ando HY: **Route designer: a retrosynthetic analysis tool utilizing automated retrosynthetic rule generation**. *J Chem Inf Model* 2009, **49**:593–602.

180. Bøgevig A, Federsel H-J, Huerta F, Hutchings MG, Kraut H, Langer T, Löw P, Oppawsky C, Rein T, Saller H: **Route design in the 21st century: the IC SYNTH software tool as an idea generator for synthesis prediction**. *Org Process Res Dev* 2015, **19**:357–368.

181. ** Coley CW, Rogers L, Green WH, Jensen KF: **Computer-Assisted retrosynthesis based on molecular similarity**. *ACS Cent Sci* 2017, **3**:1237–1245.
Single retrosynthetic steps are predicted using molecular similarity and it is shown that the reactants are in the top 10 predicted precursors 74.1% of the time. This approach does not require defining a reaction class.

182. Krebsbach D, Gelernter H, Sieburth SM: **Distributed heuristic synthesis search**. *J Chem Inf Comput Sci* 1998, **38**:595–604.

183. Badowski T, Molga K, Grzybowski BA: **Selection of cost-effective yet chemically diverse pathways from the networks of computer-generated retrosynthetic plans**. *Chem Sci* 2019, **10**:4640–4651.

184. ** Molga K, Dittwald P, Grzybowski BA: **Navigating around patented routes by preserving specific motifs along computer-planned retrosynthetic pathways**. *Chem* 2019, **5**: 460–473.
The Chematica algorithm is modified to identify alternative modes of synthesis for the blockbuster and well patent protected pharmaceuticals linezolid, sitagliptin, and panobinostat.

185. Jacob PM, Yamin P, Perez-Storey C, Hopgood M, Lapkin AA: **Towards automation of chemical process route selection based on data mining**. *Green Chem* 2017, **19**:140–152.

186. Clark PGK, Vieira LCC, Tallant C, Fedorov O, Singleton DC, Rogers CM, Monteiro OP, Bennett JM, Baronio R, Müller S, *et al.*: **LP99: discovery and synthesis of the first selective BRD7/9 bromodomain inhibitor**. *Angew Chem Int Ed* 2015, **54**: 6217–6221.

187. Wang L, Dash S, Ng CY, Maranas CD: **A review of computational tools for design and reconstruction of metabolic pathways**. *Synth Sys Biotechnol* 2017, **2**:243–252.

188. Soh KC, Hatzimanikatis V: **DREAMS of metabolism**. *Trends Biotechnol* 2010, **28**:501–508.

189. Lohr TL, Marks TJ: **Orthogonal tandem catalysis**. *Nat Chem* 2015, **7**:477.

190. Nam H, Lewis NE, Lerman JA, Lee D-H, Chang RL, Kim D, Palsson BO: **Network context and selection in the evolution to enzyme specificity**. *Science* 2012, **337**:1101.

191. Fernández-Castané A, Fehér T, Carbonell P, Pauthenier C, Faulon J-L: **Computer-aided design for metabolic engineering**. *J Biotechnol* 2014, **192**:302–313.

192. Medema MH, Van Raaphorst R, Takano E, Breitling R: **Computational tools for the synthetic design of biochemical pathways**. *Nat Rev Microbiol* 2012, **10**:191–202.

193. Kim SM, Peña MI, Moll M, Bennett GN, Kavraki LE: **A review of parameters and heuristics for guiding metabolic pathfinding**. *J Cheminf* 2017, **9**:1–13.

194. Jeffryes JG, Seaver SMD, Faria JP, Henry CS: **A pathway for every product? Tools to discover and design plant metabolism**. *Plant Sci* 2018, **273**:61–70.

195. Hatzimanikatis V, Li C, Ionita JA, Henry CS, Jankowski MD, Broadbelt LJ: **Exploring the diversity of complex metabolic networks**. *Bioinformatics* 2005, **21**:1603–1609.

196. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M: **KEGG: kyoto encyclopedia of genes and genomes**. *Nucleic Acids Res* 1999, **27**:29–34.

197. Pon A, Guo AC, Marcu A, Knox C, Grant JR, Wilson M, Ramirez M, Sajed T, Wishart DS, Djoumbou Y: **ECMDB 2.0: a richer resource for understanding the biochemistry of E. coli**. *Nucleic Acids Res* 2015, **44**:D495–D501.

198. Ravikrishnan A, Nasre M, Raman K: **Enumerating all possible biosynthetic pathways in metabolic networks**. *Sci Rep* 2018, **8**:9932.

199. Croes D, Couche F, van Helden J, Wodak SJ: **Metabolic PathFinding: inferring relevant pathways in biochemical networks**. *Nucleic Acids Res* 2005, **33**:W326–W330.

200. Faust K, Croes D, van Helden J: **Metabolic pathfinding using RPAIR annotation**. *J Mol Biol* 2009, **388**:390–414.

201. Kohlbacher O, Blum T: **MetaRoute: fast search for relevant metabolic routes for interactive network navigation and visualization**. *Bioinformatics* 2008, **24**:2108–2109.

202. Chou C-H, Chang W-C, Chiu C-M, Huang C-C, Huang H-D: **FMM: a web server for metabolic pathway reconstruction and comparative analysis**. *Nucleic Acids Res* 2009, **37**: W129–W134.

203. Küffner R, Zimmer R, Lengauer T: **Pathway analysis in metabolic databases via differential metabolic display (DMD)**. *Bioinformatics* 2000, **16**:825–836.

204. Lu W, Tamura T, Song J, Akutsu T: **Integer programming-based method for designing synthetic metabolic networks by**

minimum reaction insertion in a boolean model. *PLoS One* 2014, **9**:e92637.

205. Pey J, Prada J, Beasley JE, Planes FJ: **Path finding methods accounting for stoichiometry in metabolic networks**. *Genome Biol* 2011, **12**:R49.

206. ** Fehér T, Planson AG, Carbonell P, Fernández-Castané A, Grigoras I, Dariy E, Perret A, Faulon JL: **Validation of RetroPath, a computer-aided design tool for metabolic pathway engineering**. *Biotechnol J* 2014, **9**:1446–1457.
The Chematica algorithm is modified to identify alternative modes of synthesis for the blockbuster and well patent protected pharmaceuticals linezolid, sitagliptin, and panobinostat.

207. Liu M, Bienfait B, Sacher O, Gasteiger J, Siezen RJ, Nauta A, Geurts JMW: **Combining chemoinformatics with bioinformatics: in silico prediction of bacterial flavor-forming pathways by a chemical systems biology approach "reverse pathway engineering"**. *PLoS One* 2014, **9**:e84769.

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

209. Carbonell P, Parutto P, Herisson J, Pandit SB, Faulon J-L: **XTMS: pathway design in an eXTended metabolic space**. *Nucleic Acids Res* 2014, **42**:W389–W394.

210. * McClymont K, Soyer OS: **Metabolic tinker: an online tool for guiding the design of synthetic metabolic pathways**. *Nucleic Acids Res* 2013, **41**. e113-e113.
A thermodynamics-driven algorithm is developed to search the Universal Reaction Network for paths between any two metabolites.

211. Carbonell P, Fichera D, Pandit SB, Faulon J-L: **Enumerating metabolic pathways for the production of heterologous target chemicals in chassis organisms**. *BMC Syst Biol* 2012, **6**:10.

212. Moriya Y, Shigemizu D, Hattori M, Tokimatsu T, Kotera M, Goto S, Kanehisa M: **PathPred: an enzyme-catalyzed metabolic pathway prediction server**. *Nucleic Acids Res* 2010, **38**: W138–W143.

213. ** Delépine B, Duigou T, Carbonell P, Faulon JL: **RetroPath2.0: A retrosynthesis workflow for metabolic engineers**. *Metab Eng* 2018, **45**:158–170.
Reaction rules are encoded as SMIRK-like strings. The program provides solutions to different tasks given the rules, source compounds, and sink compounds, such as designing new pathway to a molecule and biosensor development.

214. Wicker J, Lorsbach T, Gütlein M, Schmid E, Latino D, Kramer S, Fenner K: **enviPath – the environmental contaminant biotransformation pathway resource**. *Nucleic Acids Res* 2016, **44**:D502–D508.

215. ** Kumar A, Wang L, Ng CY, Maranas CD: **Pathway design using de novo steps through uncharted biochemical spaces**. *Nat Commun* 2018, **9**. 184-184.
The rePrime and novoStoic algorithms are described for rule extraction and mass-balanced path optimization, respectively.

216. Carbonell P, Parutto P, Baudier C, Junot C, Faulon JL: **Retropath: automated pipeline for embedded metabolic circuits**. *ACS Synth Biol* 2014, **3**:565–577.

217. Sivakumar TV, Giri V, Park JH, Kim TY, Bhaduri A: **ReactPRED: a tool to predict and analyze biochemical reactions**. *Bioinformatics* 2016, **32**:3522–3524.

218. * Duigou T, du Lac M, Carbonell P, Faulon J-l: **RetroRules: a database of reaction rules for engineering biology**. *Nucleic Acids Res* 2018:1–7.
A publicly accessible database of 400,000 + reaction rules accounting for stereochemistry and provided for different distances from the reaction center.

219. Gao J, Wackett LP, Ellis LBM: **The university of Minnesota pathway prediction system: multi-level prediction and visualization**. *Nucleic Acids Res* 2011, **39**:W406–W411.

220. ** Hadadi N, Hafner J, Shajkofci A, Zisaki A, Hatzimanikatis V: **ATLAS of biochemistry: a repository of all possible biochemical reactions for synthetic biology and metabolic engineering studies**. *ACS Synth Biol* 2016, **5**:1155–1166.
A publicly available database of metabolites and reactions enumerated based on simple reaction rules.

221. Oh M, Yamada T, Hattori M, Goto S, Kanehisa M: **Systematic analysis of enzyme-catalyzed reaction patterns and prediction of microbial biodegradation pathways**. *J Chem Inf Model* 2007, **47**:1702–1712.

222. Klopman G, Dimayuga M, Talafous J: **Meta. 1. A program for the evaluation of metabolic transformation of chemicals**. *J Chem Inf Comput Sci* 1994, **34**:1320–1325.

223. Talafous J, Sayre LM, Mieyal JJ, Klopman G: **Meta. 2. A dictionary model of mammalian xenobiotic metabolism**. *J Chem Inf Comput Sci* 1994, **34**:1326–1333.

224. Hou BK, Ellis LBM, Wackett LP: **Encoding microbial metabolic logic: predicting biodegradation**. *J Ind Microbiol Biotechnol* 2004, **31**:261–272.

225. Ellis LBM, Roe D, Wackett LP: **The university of Minnesota biocatalysis/biodegradation database: the first decade**. *Nucleic Acids Res* 2006, **34**:D517–D521.

226. Stine AE, Pertusi DA, Tyo KEJ, Broadbelt LJ: **Efficient searching and annotation of metabolic networks using chemical similarity**. *Bioinformatics* 2014, **31**:1016–1024.

227. Hadadi N, Hatzimanikatis V: **Design of computational retrobiosynthesis tools for the design of de novo synthetic pathways**. *Curr Opin Chem Biol* 2015, **28**:99–104.

228. Schomburg D, Schomburg I, Leber M, Egelhofer V: **Automatic assignment of reaction operators to enzymatic reactions**. *Bioinformatics* 2009, **25**:3135–3142.

229. Cho A, Yun H, Park JH, Lee SY, Park S: **Prediction of novel synthetic pathways for the production of desired chemicals**. *BMC Syst Biol* 2010, **4**:35.

230. ** Jeffryes JG, Colastani RL, Elbadawi-Sidhu M, Kind T, Niehaus TD, Broadbelt LJ, Hanson AD, Fiehn O, Tyo KEJ, Henry CS, MINEs: **Open access databases of computationally predicted enzyme promiscuity products for untargeted metabolomics**. *J Cheminf* 2015, **7**:44.
A publicly available database of enumerated metabolic compounds and reactions, including hypothetical networks associated with specific chasses organisms.

231. Pfeiffer T, Soyer OS, Bonhoeffer S: **The evolution of connectivity in metabolic networks**. *PLoS Biol* 2005, **3**:e228.

232. Henry CS, Broadbelt LJ, Hatzimanikatis V: **Discovery and analysis of novel metabolic pathways for the biosynthesis of industrial chemicals: 3-hydroxypropanoate**. *Biotechnol Bioeng* 2010, **106**:462–473.

233. Tervo CJ, Reed JL: **MapMaker and PathTracer for tracking carbon in genome-scale metabolic models**. *Biotechnol J* 2016, **11**:648–661.

234. Huang Y, Zhong C, Lin HX, Wang J: **A method for finding metabolic pathways using atomic group tracking**. *PLoS One* 2017, **12**. e0168725-e0168725.

235. Latendresse M, Krummenacker M, Karp PD: **Optimal metabolic route search based on atom mappings**. *Bioinformatics* 2014, **30**:2043–2050.

236. Barabási A-L, Oltvai ZN: **Network biology: understanding the cell's functional organization**. *Nat Rev Genet* 2004, **5**: 101–113.

237. Pitkänen E, Jouhten P, Rousu J: **Inferring branching pathways in genome-scale metabolic networks**. *BMC Syst Biol* 2009, **3**: 1–22.

238. Noor E, Haraldsdóttir HS, Milo R, Fleming RMT: **Consistent estimation of Gibbs energy using component contributions**. *PLoS Comput Biol* 2013, **9**:e1003098.

239. Chowdhury A, Maranas CD: **Designing overall stoichiometric conversions and intervening metabolic reactions**. *Sci Rep* 2015, **5**:16009.

240. Vieira G, Portais J-C, Carnicer M, Heux S: **FindPath: a Matlab solution for in silico design of synthetic metabolic pathways**. *Bioinformatics* 2014, **30**:2986–2988.

241. ** Bar-Even A, Flamholz A, Noor E, Milo R: **eQuilibrator—the biochemical thermodynamics calculator**. *Nucleic Acids Res* 2011, **40**:D770–D775.
The authors provide a database and automated interface for estimating the Gibbs free energy of formation for biomolecules based on decades of empirical data.

242. Bar-Even A, Flamholz A, Davidi D, Noor E, Milo R, Lubling Y: **An integrated open framework for thermodynamics of reactions that combines accuracy and coverage**. *Bioinformatics* 2012, **28**:2037–2044.

243. * Noor E, Bar-Even A, Flamholz A, Reznik E, Liebermeister W, Milo R: **Pathway thermodynamics highlights kinetic obstacles in central metabolism**. *PLoS Comput Biol* 2014, **10**:e1003483.
Using the TCA cycle as an example, the authors assess the utility of thermodynamic calculations in evaluating pathway fitness.

244. Planson A-G, Carbonell P, Paillard E, Pollet N, Faulon J-L: **Compound toxicity screening and structure–activity relationship modeling in Escherichia coli**. *Biotechnol Bioeng* 2012, **109**:846–850.

245. Harder A, Escher BI, Schwarzenbach RP: **Applicability and limitation of QSARs for the toxicity of electrophilic chemicals**. *Environ Sci Technol* 2003, **37**:4955–4961.

246. Mayr A, Klambauer G, Unterthiner T, Hochreiter S: **DeepTox: toxicity prediction using deep learning**. *Front Environ Sci* 2016, **3**.

247. Rodrigo G, Jaramillo A, Carrera J, Prather KJ: **DESHARKY: automatic design of metabolic pathways for optimal cell growth**. *Bioinformatics* 2008, **24**:2554–2556.

248. Bornscheuer UT: **Biocatalysis: successfully crossing boundaries**. *Angew Chem Int Ed* 2016, **55**:4372–4373.

249. Hönig M, Sondermann P, Turner NJ, Carreira EM: **Enantioselective chemo- and biocatalysis: partners in retrosynthesis**. *Angew Chem Int Ed* 2017, **56**:8942–8973.

250. Porter JL, Rusli RA, Ollis DL: **Directed evolution of enzymes for industrial biocatalysis**. *Chembiochem* 2016, **17**:197–203.

251. Choi J-M, Han S-S, Kim H-S: **Industrial applications of enzyme biocatalysis: current status and future aspects**. *Biotechnol Adv* 2015, **33**:1443–1454.

252. France SP, Hepworth LJ, Turner NJ, Flitsch SL: **Constructing biocatalytic cascades: in vitro and in vivo approaches to de Novo multi-enzyme pathways**. *ACS Catal* 2017, **7**:710–724.

253. Singh R, Kumar M, Mittal A, Mehta PK: **Microbial enzymes: industrial progress in 21st century**. *3 Biotech* 2016, **6**:174.

254. Asano Y: **Hydrolysis of nitriles to amides**. In *Science of synthesis: biocatalysis in organic synthesis*. Georg Thieme Verlag; 2015:255–276.

255. Elisa L, Kerstin S, Anton G, Ivan H, Roger AS, Sander van Pelt, Margit W: **Mini-review: recent developments in hydroxynitrile lyases for industrial biotechnology**. *Recent Pat Biotechnol* 2013, **7**:197–206.

256. Chapman J, Ismail EA, Dinu ZC: **Industrial applications of enzymes: recent advances, techniques, and outlooks**. *Catalysts* 2018, **8**.

257. Savile CK, Janey JM, Mundorff EC, Moore JC, Tam S, Jarvis WR, Colbeck JC, Krebber A, Fleitz FJ, Brands J, *et al.*: **Biocatalytic asymmetric synthesis of sitagliptin manufacture**. *Science* 2010, **329**:305–310.

258. Zwick CR, Renata H: **Remote C-H hydroxylation by an α-ketoglutarate-dependent dioxygenase enables efficient chemoenzymatic synthesis of manzacidin C and proline analogs**. *J Am Chem Soc* 2018, **140**:1165–1169.

259. Šardzík R, Green AP, Laurent N, Both P, Fontana C, Voglmeir J, Weissenborn MJ, Haddoub R, Grassi P, Haslam SM, *et al.*: **Chemoenzymatic synthesis of O -mannosylpeptides in solution and on solid phase**. *J Am Chem Soc* 2012, **134**:4521–4524.

260. Ríos-Lombardía N, Vidal C, Liardo E, Morís F, García-Álvarez J, González-Sabín J: **From a sequential to a concurrent reaction in aqueous medium: ruthenium-catalyzed allylic alcohol isomerization and asymmetric bioreduction**. *Angew Chem Int Ed* 2016, **55**:8691–8695.

261. Köhler V, Wilson YM, Dürrenberger M, Ghislieri D, Churakova E, Quinto T, Knörr L, Häussinger D, Hollmann F, Turner NJ, *et al.*: **Synthetic cascades are enabled by combining biocatalysts with artificial metalloenzymes**. *Nat Chem* 2013, **5**:93–99.

262. Busto E, Simon RC, Kroutil W: **Vinylation of unprotected phenols using a biocatalytic system**. *Angew Chem Int Ed* 2015, **54**:10899–10902.

263. Song J-W, Jeon E-Y, Song D-H, Jang H-Y, Bornscheuer UT, Oh D-K, Park J-B: **Multistep enzymatic synthesis of long-chain α,ω-dicarboxylic and ω-hydroxycarboxylic acids from renewable fatty acids and plant oils**. *Angew Chem Int Ed* 2013, **52**:2534–2537.

264. Westfall PJ, Pitera DJ, Lenihan JR, Eng D, Woolard FX, Regentin R, Horning T, Tsuruta H, Melis DJ, Owens A, *et al.*: **Production of amorphadiene in yeast, and its conversion to dihydroartemisinic acid, precursor to the antimalarial agent artemisinin**. *Proc Natl Acad Sci Unit States Am* 2012, **109**:655.

265. Raemakers-Franken PC, Nossin PMM, Brandts PM, Wubbolts MG, Peeters WPH, Ernste S, Wildeman d Sma SM: *Biochemical synthesis of 6-amino caproic acid*. USPTO. Genomatica Inc; 2009.

266. Tsoi R, Wu F, Zhang C, Bewick S, Karig D, You L: **Metabolic division of labor in microbial systems**. *Proc Natl Acad Sci Unit States Am* 2018, **115**:2526–2531.

267. Guha TK, Wai A, Hausner G: **Programmable genome editing tools and their regulation for efficient genome engineering**. *Comput Struct Biotechnol J* 2017, **15**:146–160.

268. Liu J, Wang Y, Lu Y, Zheng P, Sun J, Ma Y: **Development of a CRISPR/Cas9 genome editing toolbox for Corynebacterium glutamicum**. *Microb Cell Factories* 2017, **16**:1–17.

269. Cook TB, Rand JM, Nurani W, Courtney DK, Liu SA, Pfleger BF: **Genetic tools for reliable gene expression and recombineering in Pseudomonas putida**. *J Ind Microbiol Biotechnol* 2018, **45**:517–527.

270. Shapiro RS, Chavez A, Collins JJ: **CRISPR-based genomic tools for the manipulation of genetically intractable microorganisms**. *Nat Rev Microbiol* 2018, **16**:333–339.

271. Cho S, Shin J, Cho B-K: **Applications of CRISPR/cas system to bacterial metabolic engineering**. *Int J Mol Sci* 2018, **19**.

272. Wang B, Hu Q, Zhang Y, Shi R, Chai X, Liu Z, Shang X, Zhang Y, Wen T: **A RecET-assisted CRISPR-Cas9 genome editing in Corynebacterium glutamicum**. *Microb Cell Factories* 2018, **17**:1–16.

273. Cobb RE, Wang Y, Zhao H: **High-Efficiency multiplex genome editing of Streptomyces species using an engineered CRISPR/cas system**. *ACS Synth Biol* 2015, **4**:723–728.

274. Nakashima N, Ohno S, Yoshikawa K, Shimizu H, Tamura T: **A vector library for silencing central carbon metabolism genes with antisense RNAs in escherichia coli**. *Appl Environ Microbiol* 2014, **80**:564–573.

275. O'Brien Edward J, Monk Jonathan M, Palsson Bernhard O: **Using genome-scale models to predict biological capabilities**. *Cell* 2015, **161**:971–987.

276. Chowdhury A, Zomorrodi AR, Maranas CD, OptForce k-: **Integrating kinetics with flux balance analysis for strain design**. *PLoS Comput Biol* 2014, **10**:e1003487.

277. Kim B, Binkley R, Kim HU, Lee SY: **Metabolic engineering of Escherichia coli for the enhanced production of l-tyrosine**. *Biotechnol Bioeng* 2018, **115**:2554–2564.

278. Long CP, Gonzalez JE, Cipolla RM, Antoniewicz MR: **Metabolism of the fast-growing bacterium Vibrio natriegens elucidated by 13C metabolic flux analysis**. *Metab Eng* 2017, **44**:191–197.

279. Hoffart E, Grenz S, Lange J, Nitschel R, Müller F, Schwentner A, Feith A, Lenfers-Lücker M, Takors R, Blombach B: **High substrate uptake rates empower Vibrio natriegens as production host for industrial biotechnology**. *Appl Environ Microbiol* 2017, **83**:1–10.

280. Dalia TN, Hayes CA, Stolyar S, Marx CJ, McKinlay JB, Dalia AB: **Multiplex genome editing by natural transformation (MuGENT) for synthetic biology in Vibrio natriegens**. *ACS Synth Biol* 2017, **6**:1650–1655.

281. Gibson DG: **Programming biological operating systems: genome design, assembly and activation**. *Nat Methods* 2014, **11**:521–526.

282. Hutchison CA, Chuang R-Y, Noskov VN, Assad-Garcia N, Deerinck TJ, Ellisman MH, Gill J, Kannan K, Karas BJ, Ma L, *et al.*: **Design and synthesis of a minimal bacterial genome**. *Science* 2016, **351**:aad6253.

283. Komatsu M, Uchiyama T, Omura S, Cane DE, Ikeda H: **Genome-minimized Streptomyces host for the heterologous expression of secondary metabolism**. *Proc Natl Acad Sci Unit States Am* 2010, **107**:2646–2651.

284. Gomez-Escribano JP, Bibb MJ: **Engineering Streptomyces coelicolor for heterologous expression of secondary metabolite gene clusters**. *Microb Biotechnol* 2011, **4**:207–215.

285. Annaluru N, Muller H, Mitchell LA, Ramalingam S, Stracquadanio G, Richardson SM, Dymond JS, Kuang Z, Scheifele LZ, Cooper EM, *et al.*: **Total synthesis of a functional designer eukaryotic chromosome**. *Science* 2014, **344**:55–58.

286. Richardson SM, Mitchell LA, Stracquadanio G, Yang K, Dymond JS, DiCarlo JE, Lee D, Huang CLV, Chandrasegaran S, Cai Y, *et al.*: **Design of a synthetic yeast genome**. *Science* 2017, **355**:1040.

287. Zaslaver A, Mayo AE, Rosenberg R, Bashkin P, Sberro H, Tsalyuk M, Surette MG, Alon U: **Just-in-time transcription program in metabolic pathways**. *Nat Genet* 2004, **36**:486.

288. Staunton J, Weissman KJ: **Polyketide biosynthesis: a millennium review**. *Nat Prod Rep* 2001, **18**:380–416.

289. Wong FT, Khosla C: **Combinatorial biosynthesis of polyketides—a perspective**. *Curr Opin Chem Biol* 2012, **16**:117–123.

290. Khosla C, Herschlag D, Cane DE, Walsh CT: **Assembly line polyketide synthases: mechanistic insights and unsolved problems**. *Biochemistry* 2014, **53**:2875–2883.

291. Klaus M, Grininger M: **Engineering strategies for rational polyketide synthase design**. *Nat Prod Rep* 2018, **35**:1070–1081.

292. Wolański M, Łebkowski T, Kois-Ostrowska A, Zettler J, Apel AK, Jakimowicz D, Zakrzewska-Czerwińska J: **Two transcription factors, CabA and CabR, are independently involved in multilevel regulation of the biosynthetic gene cluster encoding the novel aminocoumarin, cacibiocin**. *Appl Microbiol Biotechnol* 2016, **100**:3147–3164.

293. Tahlan K, Ahn SK, Sing A, Bodnaruk TD, Willems AR, Davidson AR, Nodwell JR: **Initiation of actinorhodin export in Streptomyces coelicolor**. *Mol Microbiol* 2007, **63**:951–961.

294. Wei J, Tian Y, Niu G, Tan H, GouR: **A TetR family transcriptional regulator, coordinates the biosynthesis and export of gougerotin in Streptomyces graminearus**. *Appl Environ Microbiol* 2014, **80**:714–722.

295. Horbal L, Rebets Y, Rabyk M, Luzhetskyy A, Ostash B, Welle E, Nakamura T, Fedorenko V, Bechthold A: **Characterization and analysis of the regulatory network involved in control of lipomycin biosynthesis in Streptomyces aureofaciens Tü117**. *Appl Microbiol Biotechnol* 2010, **85**:1069–1079.

296. Park SY, Yang D, Ha SH, Lee SY: **Metabolic engineering of microorganisms for the production of natural compounds**. *Adv Biosys* 2018, **2**:1700190.

297. Moser F, Espah Borujeni A, Ghodasara AN, Cameron E, Park Y, Voigt CA: **Dynamic control of endogenous metabolism with combinatorial logic circuits**. *Mol Syst Biol* 2018, **14**. e8605-e8605.

298. Zhang F, Carothers JM, Keasling JD: **Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids**. *Nat Biotechnol* 2012, **30**:354–359.

299. Dahl RH, Zhang F, Alonso-Gutierrez J, Baidoo E, Batth TS, Redding-Johanson AM, Petzold CJ, Mukhopadhyay A, Lee TS, Adams PD, *et al.*: **Engineering dynamic pathway regulation using stress-response promoters**. *Nat Biotechnol* 2013, **31**:1039–1046.

300. Doong SJ, Gupta A, Prather KLJ: **Layered dynamic regulation for improving metabolic pathway productivity in Escherichia coli**. *Proc Natl Acad Sci Unit States Am* 2018, **115**:2964–2969.

301. Venayak N, Anesiadis N, Cluett WR, Mahadevan R: **Engineering metabolism through dynamic control**. *Curr Opin Biotechnol* 2015, **34**:142–152.

302. Andrews LB, Nielsen AAK, Voigt CA: **Cellular checkpoint control using programmable sequential logic**. *Science* 2018, **361**:eaap8987.

303. Endalur Gopinarayanan V, Nair NU: **A semi-synthetic regulon enables rapid growth of yeast on xylose**. *Nat Commun* 2018, **9**:1233.
To overcome growth limitations by yeast on xylose, they create alternative Gal4 regulatory circuits that do not trigger the starvation response.

304. * Schmitt DL: **An S: spatial organization of metabolic enzyme complexes in cells**. *Biochemistry* 2017, **56**:3184–3196.

305. Weissman KJ, Leadlay PF: **Combinatorial biosynthesis of reduced polyketides**. *Nat Rev Microbiol* 2005, **3**:925.

306. Wilkinson B, Kendrew SG, Sheridan RM, Leadlay PF: **Biosynthetic engineering of polyketide synthases**. *Expert Opin Ther Pat* 2003, **13**:1579–1606.

307. Crawford JM, Thomas PM, Scheerer JR, Vagstad AL, Kelleher NL, Townsend CA: **Deconstruction of iterative multidomain polyketide synthase function**. *Science* 2008, **320**:243–246.

308. Dutta S, Whicher JR, Hansen DA, Hale WA, Chemler JA, Congdon GR, Narayan ARH, Håkansson K, Sherman DH, Smith JL, *et al.*: **Structure of a modular polyketide synthase**. *Nature* 2014, **510**:512.

309. Tsai S-C, Miercke LJW, Krucinski J, Gokhale R, Chen JC-H, Foster PG, Cane DE, Khosla C, Stroud RM: **Crystal structure of the macrocycle-forming thioesterase domain of the erythromycin polyketide synthase: versatility from a unique substrate channel**. *Proc Natl Acad Sci Unit States Am* 2001, **98**:14808–14813.

310. Buchholz TJ, Geders TW, Bartley FE, Reynolds KA, Smith JL, Sherman DH: **Structural basis for binding specificity between subclasses of modular polyketide synthase docking domains**. *ACS Chem Biol* 2009, **4**:41–52.

311. Weissman KJ, Müller R: **Protein–protein interactions in multienzyme megasynthetases**. *Chembiochem* 2008, **9**:826–848.

312. Straight PD, Fischbach MA, Walsh CT, Rudner DZ, Kolter R: **A singular enzymatic megacomplex from Bacillus subtilis**. *Proc Natl Acad Sci Unit States Am* 2007, **104**:305–310.

313. Imperi F, Visca P: **Subcellular localization of the pyoverdine biogenesis machinery of Pseudomonas aeruginosa: a membrane-associated "siderosome"**. *FEBS (Fed Eur Biochem Soc) Lett* 2013, **587**:3387–3391.

314. Lee H, DeLoache WC, Dueber JE: **Spatial organization of enzymes for metabolic engineering**. *Metab Eng* 2012, **14**:242–251.

315. Li H, Zheng G, Zhu S: **Construction of an organelle-like nanodevice via supramolecular self-assembly for robust biocatalysts**. *Microb Cell Factories* 2018, **17**:26.

316. Wu GC, Anderluh G, Xu H, Vovk I, Boock JT, Turnšek J, Mori J, Dueber JE, Benčina M, Avbelj M, *et al.*: **DNA-guided assembly of biosynthetic pathways promotes improved catalytic efficiency**. *Nucleic Acids Res* 2011, **40**:1879–1889.

317. * Sachdeva G, Garg A, Godding D, Way JC, Silver PA: **In vivo co-localization of enzymes on RNA scaffolds increases metabolic production in a geometrically dependent manner**. *Nucleic Acids Res* 2014, **42**:9493–9503.
Enzymes fused with RNA binding domains were brought together via RNA scaffolds *in vivo*, resulting in the increased production of pentadecane and succinate.

318. Siu K-H, Chen RP, Sun Q, Chen L, Tsai S-L, Chen W: **Synthetic scaffolds for pathway enhancement**. *Curr Opin Biotechnol* 2015, **36**:98–106.

319. Whitaker WR, Dueber JE: **Chapter nineteen – metabolic pathway flux enhancement by synthetic protein scaffolding**. In *Methods in enzymology*. Edited by Voigt C, Academic Press; 2011:447–468.

320. Myhrvold C, Polka JK, Silver PA: **Synthetic lipid-containing scaffolds enhance production by colocalizing enzymes**. *ACS Synth Biol* 2016, **5**:1396–1403.

321. Moon TS, Dueber JE, Shiue E, Prather KLJ: **Use of modular, synthetic scaffolds for improved production of glucaric acid in engineered E. coli**. *Metab Eng* 2010, **12**:298–305.

322. Avalos JL, Fink GR, Stephanopoulos G: **Compartmentalization of metabolic pathways in yeast mitochondria improves the production of branched-chain alcohols**. *Nat Biotechnol* 2013, **31**:335.

323. Chau Angela H, Walter Jessica M, Gerardin J, Tang C, Lim Wendell A: **Designing synthetic regulatory networks capable of self-organizing cell polarization**. *Cell* 2012, **151**:320–332.

324. Reinkemeier CD, Girona GE, Lemke EA: **Designer membraneless organelles enable codon reassignment of selected mRNAs in eukaryotes**. *Science* 2019, **363**. eaaw2644.

325. Chowdhury C, Sinha S, Chun S, Yeates TO, Bobik TA: **Diverse bacterial microcompartment organelles**. *Microbiol Mol Biol Rev* 2014, **78**:438–468.

326. Kerfeld CA, Heinhorst S, Cannon GC: **Bacterial microcompartments**. *Annu Rev Microbiol* 2010, **64**:391–408.

327. Jakobson CM, Tullman-Ercek D, Slininger MF, Mangan NM: **A systems-level model reveals that 1,2-Propanediol utilization microcompartments enhance pathway flux through intermediate sequestration**. *PLoS Comput Biol* 2017, **13**: e1005525.

328. Jakobson CM, Slininger Lee MF, Tullman-Ercek D: **De novo design of signal sequences to localize cargo to the 1,2-propanediol utilization microcompartment**. *Protein Sci* 2017, **26**:1086–1092.

329. Jakobson CM, Chen Y, Slininger MF, Valdivia E, Kim EY, Tullman-Ercek D: **Tuning the catalytic activity of subcellular nanoreactors**. *J Mol Biol* 2016, **428**:2989–2996.

330. Lee MJ, Brown IR, Juodeikis R, Frank S, Warren MJ: **Employing bacterial microcompartment technology to engineer a shell-free enzyme-aggregate for enhanced 1,2-propanediol production in Escherichia coli**. *Metab Eng* 2016, **36**:48–56.

331. Held M, Kolb A, Perdue S, Hsu S-Y, Bloch SE, Quin MB, Schmidt-Dannert C: **Engineering formation of multiple recombinant Eut protein nanocompartments in E. coli**. *Sci Rep* 2016, **6**: 24359.

332. Chessher A, Breitling R, Takano E: **Bacterial microcompartments: biomaterials for synthetic biology-based compartmentalization strategies**. *ACS Biomater Sci Eng* 2015, **1**: 345–351.

333. Kim EY, Tullman-Ercek D: **Engineering nanoscale protein compartments for synthetic organelles**. *Curr Opin Biotechnol* 2013, **24**:627–632.

334. Glasgow JE, Asensio MA, Jakobson CM, Francis MB, Tullman-Ercek D: **Influence of electrostatics on small molecule flux through a protein nanoreactor**. *ACS Synth Biol* 2015, **4**: 1011–1019.

335. Chowdhury C, Chun S, Pang A, Sawaya MR, Sinha S, Yeates TO, Bobik TA: **Selective molecular transport through the protein shell of a bacterial microcompartment organelle**. *Proc Natl Acad Sci Unit States Am* 2015, **112**:2990–2995.

336. ** Lau YH, Giessen TW, Altenburg WJ, Silver PA: **Prokaryotic nanocompartments form synthetic organelles in a eukaryote**. *Nat Commun* 2018, **9**:1311.
Self-assembled encapsulin are used to encapsulate proteins to prevent proteins from protelyotic degration, co-localize enzymes, and protect unstable or toxic enzymatic intermediates.

337. Hinzpeter F, Gerland U, Tostevin F: **Optimal compartmentalization strategies for metabolic microcompartments**. *Biophys J* 2017, **112**:767–779.

338. Jakobson CM, Tullman-Ercek D, Mangan NM: **Spatially organizing biochemistry: choosing a strategy to translate synthetic biology to the factory**. *Sci Rep* 2018, **8**:8196.

339. Liu Q, Manzano D, Tanić N, Pesic M, Bankovic J, Pateraki I, Ricard L, Ferrer A, de Vos R, de Krol Sv, *et al.*: **Elucidation and in planta reconstitution of the parthenolide biosynthetic pathway**. *Metab Eng* 2014, **23**:145–153.

340. Yang Z-J, Ge W-Z, Li Q-Y, Lu Y, Gong J-M, Kuang B-J, Xi X, Wu H, Zhang Q, Chen Y: **Syntheses and biological evaluation of costunolide, parthenolide, and their fluorinated analogues**. *J Med Chem* 2015, **58**:7007–7020.

341. Nakamura CE, Whited GM: **Metabolic engineering for the microbial production of 1,3-propanediol**. *Curr Opin Biotechnol* 2003, **14**:454–459.

342. Paddon CJ, Keasling JD: **Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development**. *Nat Rev Microbiol* 2014, **12**:355.

343. Buijs NA, Siewers V, Nielsen J: **Advanced biofuel production by the yeast Saccharomyces cerevisiae**. *Curr Opin Chem Biol* 2013, **17**:480–488.

344. Meadows AL, Hawkins KM, Tsegaye Y, Antipov E, Kim Y, Raetz L, Dahl RH, Tai A, Mahatdejkul-Meadows T, Xu L, *et al.*: **Rewriting yeast central carbon metabolism for industrial isoprenoid production**. *Nature* 2016, **537**:694.

345. Coordinators NR: **Database resources of the national center for biotechnology information**. *Nucleic Acids Res* 2017, **46**: D8–D13.

346. Jiang G-Z, Yao M-D, Wang Y, Zhou L, Song T-Q, Liu H, Xiao W-H, Yuan Y-J: **Manipulation of GES and ERG20 for geraniol overproduction in Saccharomyces cerevisiae**. *Metab Eng* 2017, **41**:57–66.

347. Synthetic Reaction Updates.http://pubs.rsc.org/lus/synthetic-reaction-updates.

348. SPRESI[web] https://www.spresi.com/.

349. SMIRKS: *A reaction transform language*. Daylight Chemical Information Systems, Inc.; 2007. http://www.daylight.com/dayhtml/doc/theory/theory.smirks.html.

350. Gao J, Wackett LP, Ellis LBM: **The University of Minnesota Biocatalysis/Biodegradation Database: improving public access**. *Nucleic Acids Res* 2009, **38**:D488–D491.

351. Kumar A, Suthers PF, Maranas CD: **MetRxn: a knowledgebase of metabolites and reactions spanning metabolic models and databases**. *BMC Bioinf* 2012, **13**:6.

352. Moretti S, Martin O, Van Du T, Bridge A, Morgat A, Pagni M: **MetaNetX/MNXref - reconciliation of metabolites and biochemical reactions to bring together genome-scale metabolic networks**. *Nucleic Acids Res* 2016, **44**:D523–D526.