Biosynthesis of Therapeutic Natural Products using Synthetic Biology

Ali R. Awan[a,b], William M. Shaw[a,b], Thomas Ellis[a,b*]

[a] Centre for Synthetic Biology and Innovation, Imperial College London, London SW7 2AZ

[b] Department of Bioengineering, Imperial College London, London SW7 2AZ

* corresponding author: t.ellis@imperial.ac.uk

Abstract

Natural products are a group of bioactive structurally diverse chemicals produced by microorganisms and plants. These molecules and their derivatives have contributed to over a third of the therapeutic drugs produced in the last century. However over the last few decades traditional drug discovery pipelines from natural products have become far less productive and far more expensive. One recent development with promise to combat this trend is the application of synthetic biology to therapeutic natural product biosynthesis. Synthetic biology is a young discipline with roots in systems biology, genetic engineering and metabolic engineering. In this review we discuss the use of synthetic biology to engineer improved yields of existing therapeutic natural products. We further describe the use of synthetic biology to combine and express natural product biosynthetic genes in unprecedented ways, and how this holds promise for opening up completely new avenues for drug discovery and production.

Contents

1. Introduction

1.1 Natural Products as Therapeutics

Chemicals found in nature have been used for therapeutic purposes since ancient times. The ancient Egyptians and Greeks used salicylic acid from the bark of the willow tree to treat aches and pains [1]. In the 19th century chemists at Bayer modified this substance to make aspirin. Natural products like salicylic acid are small molecules produced by plants, bacteria and fungi that have been selected by evolution for stability and interaction with biological polymers (proteins, nucleic acids, carbohydrates and lipid membranes) [2]. Their importance to human health is underscored by the fact that natural products and their derivatives comprise over 40% of drugs, including antibiotics and anti-tumour and cholesterol-lowering agents [3, 4]. The major classes of therapeutic natural product, along with an illustrative member of the class and the organism it was discovered in is given in figure 1.

One class of therapeutic drug in particular, the antibiotics, has relied heavily on natural products. The discovery of penicillin in 1928 heralded the “Golden Era” of natural products as antibiotics [5]. In the decades since the 1980s, however, fewer and fewer new antibiotics were being discovered in nature [6]. This situation led to a prioritising of pharmaceutical drug discovery towards completely synthetic chemical avenues [7]. However, several recent trends have motivated a move back to exploiting natural products [2]. This review focuses on one of these developments: the application of synthetic biology towards the production of natural product-derived therapeutic drugs.

1.2 Synthetic Biology for Therapeutic Production

Synthetic biology has several definitions. Here we adhere to the definition that the goal of synthetic biology is to extend or modify the behaviour of organisms using molecular biology to perform new tasks in a predictable manner [8]. The relevant task in this case is the production of therapeutic natural products at commercially viable yields by a suitable host organism.

Natural products are mostly produced by the action of multiple genes [9]. In the simplest scenario each gene encodes an enzyme that converts an input chemical into an output chemical acted upon by the next enzyme until the final natural product is produced, in assembly-line fashion. Together, these biosynthetic genes comprise a pathway. To produce natural products at high yields, a synthetic biologist must balance pathway gene expression and host cell growth. This endeavour must acknowledge the burden caused by the pathway gene expression, both via siphoning away of host resources and via the build-up of toxic pathway intermediate products [10]. In order to achieve this balance and maximise pathway yield, synthetic biology has developed tools to allow finely-tuned control over pathway behaviour.

These tools fall into two main categories. The first category of tools has roots in the older disciplines of genetic engineering and metabolic engineering, and includes simple mutation and screening, rational modulation of host organism gene expression, protein engineering, directed evolution, and optimisation of growth conditions [11]. The second class of tools can be viewed as purely belonging to synthetic biology. These tools are based on adapting naturally occurring biological molecules at the DNA, RNA and protein levels to confer the desired behaviour to pathway function [12].

When it comes to optimising the production of natural products, the distinction between the terms “metabolic engineering” and “synthetic biology” is breaking down, as these terms are increasingly being used interchangeably in the literature [13, 14]. Thus this review discusses both categories of tools mentioned above, but specifically using cases concerning the production of natural products with therapeutic value.

2. Known Natural Products

In the majority of cases taken on by synthetic biology, the chemical structure of the natural product is known. The main tasks faced by synthetic biologists in these cases are threefold. The genes encoding enzymes that will convert starting chemicals into the final natural product must be selected, and a host organism must be chosen. Finally, ways must be found to control the expression of these genes to optimally balance pathway yield and host organism growth. These tasks are not independent, for example, certain paradigms for gene expression control are specific to prokaryotes or eukaryotes, and the cellular conditions of different hosts may favour different types of enzymatic reaction, necessitating different pathway genes [11]. This process can be illustrated by many recent examples and especially by the most celebrated case of a natural product made using synthetic biology – that of the anti-malarial drug precursor, artemisinic acid [10, 15].

2.1 Choosing the Host Organism and Enzymatic Steps

2.1.1 Natural Hosts

The simplest cases involve optimising the production of molecules already made by an organism. A natural host is desirable when aspects of the cell biology and metabolism of the native host render them optimal for the production of certain compounds. Indeed, for perhaps the most prolific genus of natural product producers, the bacterial genus *Streptomyces*, this is precisely the case [2]. Certain biosynthetic enzymes responsible for the production of natural products in *Streptomyces* do not function well in common heterologous host organisms [16]. A similarly prolific

natural producer, the fungal genus *Aspergillus*, is also considered a good host for biosynthesis of natural products that originate within its genus [17].

Species of both *Streptomyces* and *Aspergillus* produce a plethora of natural products, so when the objective is to produce a single natural product at commercially viable levels, one approach has been to delete or inactivate competing natural product pathways [18]. Further advances have involved developing synthetic gene regulatory elements. One prominent example is promoters, elements responsible for the strength of gene transcription. Synthetic promoters with known strengths and promoters that are conditionally regulated have been developed to control pathway expression in natural *Streptomyces* and *Aspergillus* hosts [19, 20]. This process of creating a rationally optimised "natural" host strain allowed researchers to harness the natural production capabilities of organisms that make native therapeutic natural products, to make heterologous natural products. This endeavour can be viewed as an intermediate case between using a completely natural host, and using a heterologous host as described below.

2.1.2 Heterologous Hosts

The metabolic cost of producing natural products is such that they are usually produced at levels too low for commercial viability [11]. Thus the production needs to be optimised by a combination of genetic manipulation and optimisation of growth conditions. However in most cases, natural products are usually only found in organisms that are either not amenable to optimisation by genetic means or are unsuitable for growth in the large-scale industrial vessels required for high titres [21]. In other cases the desired pathway uses genes from many different organisms, so by definition there can be no completely natural host [22]. In these scenarios using a heterologous host provides a solution.

For industrial production of natural products, the most widely used heterologous host organisms are the bacterium *Escherichia coli* and the yeast *Saccharomyces cerevisiae*. The reasons include their ease of genetic manipulation, their extremely well-understood cell biology, the availability of extensive metabolic models, and their ability to tolerate industrial growth conditions. Other important heterologous hosts include the yeast *Pichia pastoris,* which can be grown on cheap carbon sources*,* and the fungus *Aspergillus oryzae*, which has cell biology suitable for the production of polyketides and nonribosomal peptides [23].

An important consideration for the choice of host is whether there is pathway-specific biology that cannot be achieved by every host, although these issues can sometimes be solved by genetic engineering. For example, some natural products require certain enzymatic reactions to occur in sub-cellular organelles. For instance, the final reactions in penicillin biosynthesis occur naturally in peroxisomes [24]. Recent work to produce penicillin in a heterologous host used the yeast *Hansenula polymorpha* due to its extremely large peroxisomes [25, 26]. Another salient example concerns the expression of cytochrome p450 genes. These genes are extensively

involved in the biosynthesis of therapeutic natural products such as artemisinic acid, but their expression in *E. coli*, has been problematic [27]. Cytochromes p450 are not naturally found in *E. coli*, and that organism lacks some of the cellular machinery for their proper expression [28]. By contrast, yeasts such as *S. cerevisiae* and *P. pastoris* have naturally occurring cytochrome p450s and might represent better hosts for certain cytochrome p450 containing pathways [29]. For example, a recent study simultaneously expressed multiple cytochromes p450 in *S. cerevisiae* to make ginsenosides, potentially therapeutic natural products that are the primary bioactive compounds of ginseng [30].

2.1.3 Multiple Hosts

Sometimes no single organism can adequately carry out the entire biosynthetic pathway for a given natural product. However, pathways can sometimes be split up between different host organisms. In a recent study from Zhou and co-workers, the plant natural product paclitaxel, an anti-tumour drug, was produced by splitting the biosynthetic pathway between the model heterologous hosts *E. coli* and *S. cerevisiae [31]*.

The paclitaxel pathway can be conceptualised as a preliminary stage producing the precursor taxadiene, and a second stage chemically functionalising the taxadiene via oxygenation reactions. Production of the precursor taxadiene at high yields was hypothesised to be more efficiently carried out by *E. coli* due to its faster growth dynamics relative to *S. cerevisiae*. Due to its cellular biology and the presence of abundant cellular membranes, *S. cerevisiae* was hypothesised as a good host for the second part of the pathway, mediating the oxygenation reactions. Indeed, when these two microorganisms were engineered to harbour the genes encoding the respective parts of the pathway and then co-cultured, the resulting consortium produced 33 mg/L of paclitaxel. A key point was that the authors used the different biology of the two host organisms to ensure co-operation to make the natural product. Namely, *S. cerevisiae* was unable to use xylose in the media as a carbon source, but could use the acetate produced by *E. coli*, as shown in Figure 2b. When they tried a similar strategy using two different engineered strains of *E. coli*, the yield was much lower.

2.1.4 Choosing Enzymatic Steps for the Pathway

For a given natural product there is often more than one particular biosynthetic route of enzymatic steps that take inputs to make the final product [23]. Depending on the host organism chosen, more or less of the pathway can be carried out by native genes in the host without relying on heterologous genes. This concept can be illustrated by the biosynthesis of artemisinic acid, precursor to the anti-malarial drug artemisinin. The first part of the pathway requires the production of the precursor farnesyl pyrophosphate (FPP). This precursor is produced natively by the primary

metabolism of both *S. cerevisiae* and *E. coli*. In the original work to produce artemisinic acid, Keasling and co-workers used *S. cerevisiae* yeast as a host organism, and were able to recruit the *S. cerevisiae* mevalonate pathway to produce the required FPP [15]. Concurrent work to produce artemisinic acid in *E. coli* could have utilised the native *E. coli* deoxyxyulos-5-phosphate pathway to produce FPP, but instead chose to import the mevalonate pathway from *S. cerevisiae [10]*. These cases illustrate both how there is often a choice of which enzymatic steps to take from metabolic inputs to natural product, and how that choice can depend on the particular host organism.

2.2 Optimising Pathway Yield

After initial expression of the pathway genes in the chosen host and successful production of the desired product, the next set of tasks involve increasing the yield to commercially viable levels. Simply over-expressing every pathway gene is usually insufficient, as there is a burden on the host caused by heterologous gene expression, and stresses also arise from any protein mis-folding and aggregation [32]. Foreign enzymes usually also cause further 'metabolic burden' by siphoning away key metabolites and co-factors from host primary metabolism, consuming energy during their reactions (e.g. via ATP use) and by altering the redox state of the cell. Furthermore, many enzymes within a pathway will also produce intermediates that are toxic to the host organism [33, 34]. Therefore, rather than over-expressing each enzyme in the pathway, expressing just enough to allow for efficient catalytic turnover between metabolic inputs and intermediates is often a better approach [10]. The most important approaches developed by synthetic biology to achieve balance between pathway yield and host health are discussed below.

2.2.1 Rational Approaches

Rational approaches involve making specific, calculated changes to the production system that are predicted to allow more throughput, or "flux", through the pathway from while not increasing burden on the host. These changes are predicted based on metabolic modelling methods that rely on measuring metabolic flux through the pathway. These measurements typically come from experimental techniques such as LCMS, GCMS and proteomics, combined with computational methods that take into account genome-scale metabolic models. For more detail the reader is directed to several excellent reviews covering these methods [35-38].

2.2.1.1 Increasing Precursor Supply and Inactivating Competing Pathways

One conceptually straightforward means of increasing the yield of a natural product pathway is to genetically modify the host organism to bias its primary metabolism to produce more of a pathway

246 substrate, as illustrated in Figure 2 [11]. The early work on artemisinic acid did exactly this, by
247 increasing the expression levels of modified versions of the *S. cerevisiae* mevalonate genes
248 *HGMR* and *ERG20 [15]*. These are both host genes responsible for the production of FPP, a
249 precursor to artemisinic acid. Over-expression was achieved by integrating additional copies of
250 both genes into the host genome. Aside from up-regulating the expression of genes whose activity
251 directly increases the amounts of a precursor, another strategy shown in Figure 2 is to inactivate
252 host metabolic pathways that siphon away the precursor for competing products [11]. In the case
253 of artemisinin production, the *S. cerevisiae ERG9* gene was down-regulated by placing it under the
254 control of a methionine-repressible promoter. *ERG9* deletion was not an option, since it is a gene
255 essential for growth of the host organism. However, inducible down-regulation was able to
256 decrease the amount of the artemisinin precursor FPP being consumed in the host's competing
257 sterol biosynthesis pathway.
258 More recently, pioneering work from several different labs has made feasible the production
259 of the medically important benzylisoquinoline alkaloids (BIAs) class of natural products so that
260 these can be made from simple sugars by *S. cerevisiae* yeast [39]. Prominent examples of BIAs
261 include reticuline, morphine and codeine. In one of the key studies enabling this production,
262 Trenchard and co-workers increased the supply of the BIA precursor tyrosine to achieve a
263 increased production of norcoclaurine, a BIA scaffold molecule [40]. Highlighting the power of
264 using a well-studied host organism, this was done by identifying six genetic modifications in *S.*
265 *cerevisiae* yeast that genome-scale metabolic models predicted by genome-scale metabolic
266 modelling to increase the supply of intracellular tyrosine. Three of the six postulated changes
267 proved to be effective, and combined gave a 60-fold increase in the levels of the norcoclaurine
268 produced.
269
270 *2.2.1.2 Spatial Strategies*
271
272 An important concept in optimising yields of a heterologous pathway is the efficient channelling of
273 metabolic intermediates. Channelling entails quickly co-ordinating the products of one enzyme in a
274 pathway to serve as a substrate the next enzyme in the pathway. This gives two advantages for
275 balancing product yield and host organism health. First, there is less time for competing pathways
276 to consume any intermediates if they are processed quickly through the pathway. Second, this
277 prevents the build-up of toxic pathway intermediates within the cell.
278 The obvious strategy, and one harnessed by nature in some native metabolic pathways
279 [34], is spatial organisation of the pathway enzymes in a way that encourages the efficient
280 channelling of intermediates. One synthetic biology mechanism to achieve this is scaffolding,
281 where pathway enzymes are tethered to a synthetic physical scaffold in a manner that places
282 active sites in coordination

The earliest demonstration of scaffolding in natural product biosynthesis was achieved by Dueber *et al* who built several versions of a synthetic protein scaffold system by interspersing a varying number of the protein binding domains GBD, SH3 and PDZ with linker regions consisting of non-structure forming amino acids [41]. They then attached the cognate partners of each of these domains to the C-terminals of the three enzymes that produce mevalonate (a precursor to the artemisinin precursor, FPP), AtoB, HMGS and HMGR respectively, as shown in Figure 2b. These tagged enzymes now bound to the synthetic protein scaffolds and the best combination of one AtoB, two HMGSs and two HMGRs enzymes aligned on a scaffold gave a 77-fold improvement in yield over simple expression of the free enzymes. A useful aspect of scaffolding is that the relative stoichiometry of pathway enzymes can easily be titrated. Aside from protein scaffolds, RNA and DNA-based scaffolds have also been used to increase pathway yield in other studies [42, 43].

The second spatial strategy is compartmentalisation of pathway enzymes, shown in Figure 2b. In nature, eukaryotic cells confine certain biochemical pathways to various organelles and prokaryotic cells to various proteinacious compartments [44]. Compartmentalisation not only increases local enzyme concentration and prevents build-up of toxic intermediates but further protects the host organism from any toxic metabolites, and can provide specialised conditions for enzymatic reactions, such as a different pH or altered salt or input metabolite concentrations [45, 46]. Therapeutic natural product pathways that use compartmentalisation include the production of the antibiotic bacillaene in a membrane-bound subcellular compartment of *B. subtilis* [47], and the occurrence of the last two steps of penicillin biosynthesis in the peroxisomes of filamentous fungi [34]. Compartmentalisation has also been used as a strategy for the manufacture of a natural product in a heterologous host. An important early example concerns the production of plant terpenoids in *S. cerevisiae*. The engineering of plant sesquiterpene production into heterologous hosts has already been discussed with regard to the therapeutic natural products artemisinin and paclitaxel. In this study the production of a precursor to artemisinin – amorphadiene – was optimised by compartmentalisation. Specifically, researchers targeted the final two enzymes of the amorphadiene biosynthesis pathway to the mitochondria of *S. cerevisiae*. In doing so they achieved an increase in amorphadiene yield of 20-fold over a version of the pathway in which these enzymes were not compartmentalised [48].

2.2.1.3 Dynamic Control

The above approaches all involve changing gene expression or enzyme localisation in static ways. However, one area where synthetic biology has made great progress is in engineering organisms to perform logical operations [49]. This involves building systems able to produce different outputs given different inputs. An important example of how this aids natural product manufacturing is via inducible gene expression. Ever since mathematical logic was first discovered to apply to gene

regulation [50] biology has utilised promoters that can induce or repress gene expression based on the presence or absence of small molecules such as arabinose and tetracycline. Efforts in synthetic biology have vastly expanded the number of available promoters responsive to such molecules [51, 52].

Inducible promoters have been important for biosynthesis as they allow production by a pathway to be switched on when desired. This is usually only after the host organism has reached exponential growth phase, allowing host resources to be used for growth initially, and then diverted to biosynthesis when enough host cells are present and a stable maintenance phase is reached [23, 53]. The construction of synthetic promoters induced by a variety of molecules is an important endeavour that enables improved metabolic engineering and it has now been extended to a variety of host organisms. In the important natural producer genus *Streptomyces*, several promoters have been constructed that are inducible by small molecules such as thiostrepton [54], gamma-butyrolactones [55] and tetracycline [56]. There are also now tetracycline-inducible promoters available for work in cyanobacteria [57]. In more established hosts, important new inducible promoters have also been made. Aromatic amino acid-inducible promoters constructed for use *S. cerevisiae [58]* now enable expression of pathway enzymes to be responsive to molecules that are themselves precursors or intermediates in natural product pathways.

In order to extend the paradigm of inducible or repressible gene expression, synthetic biology has also exploited RNA elements such as riboswitches. These are RNA regions whose three dimensional conformation allows it to bind to a small molecule and then change shape in a manner that influence the surrounding RNA. Natural riboswitches are usually found in the non-coding regions of messenger RNAs (*i.e.* the 5'UTR, introns, 3'UTR), and binding of small molecules by riboswitches influence the expression from mRNA either positively and negatively via regulation of transcription, translation or mRNA processing [59]. The part of the riboswitch that binds the small molecule is called the "aptamer domain", and the part mediating changes in gene expression is called the "expression domain". Synthetic biology has recently discovered rules for the engineering of both domains [60], so that riboswitches can now to some extent be rationally designed using computational approaches [61]. One promising recent example by Amin and co-workers uses physics-based modelling of RNA-folding to better design riboswitches that affect translation in response to various ligands [62]. This approach achieved some success in designing riboswitches that displayed improved modularity, an important attribute for synthetic biology applications.

2.2.2 Semi-rational approaches

The approaches to pathway expression control discussed so far all seek to apply specific, predetermined changes to the system, usually informed by measurement and modelling studies. These changes are informed by the use of experimental approaches such as LCMS and

proteomics in conjunction with genome-scale metabolic models and computational algorithms. There is another class of tools useful when it is not known what changes will bring about the desired pathway behaviour. These tools instead seek to explore various solution spaces in a more or less targeted fashion. Another important aspect of optimising pathway yield in a semi-rational manner concerns optimising growth media for the producer organism(s). For example, consideration must be given to the temperature and pH chosen for production. Often these parameters are explored in a statistical manner that efficiently explores the media composition space, such as “design of experiments” [63].

2.2.2.1 Exploring Pathway Gene Expression

In natural product pathways and indeed biosynthetic pathways in general, the levels of each of the pathway enzymes relative to one another have been honed by evolution for an optimum balance between pathway expression and host fitness [11]. When natural product pathways are engineered into new hosts, it is most often not clear *a priori* what the optimum relative expression levels for natural product pathway genes are. As such, it is useful to explore the ‘expression space’ by testing different levels of expression for each pathway gene. An area in which synthetic biology has devoted much effort has been the development of part libraries for expression control. In engineering parlance, a part is a standalone object whose properties with regard to the systems in which it will be used are well understood in any context. In practical terms this ideal is much more a reality in other engineering disciplines than in synthetic biology where much is still unpredictable about how parts will behave in different contexts [64]. Nonetheless, great strides have been made to develop libraries of gene regulatory parts for different hosts, such as promoters and transcription terminators (for eukaryotic and prokaryotic hosts) and ribosome binding site (RBS) sequences (for prokaryotic hosts).

Synthetic biology kits have been developed to use these part libraries in conjunction with recently developed combinatorial DNA cloning techniques [65, 66] to allow the rapid assembly of genes and pathways in a modular fashion. These kits typically contain libraries of parts that direct different strengths of enzyme expression, and are particularly of use for the standard hosts *E. coli* [67] and *S. cerevisiae* [68]. To illustrate the value of modularity in building a pathway, we can consider a hypothetical natural product pathway with three genes to be expressed in *E. coli*. Using a DNA assembly kit with part libraries we can assemble the protein coding regions of the three genes with each type of regulatory part (promoter, ribosome binding site and terminator). If the kit contains three versions of each of these parts, each directing different strengths of enzyme expression, then this gives over 200 different combinations that could be built to make the pathway and each of these will yield different levels of the three enzymes within the cell. With modern DNA assembly methods, all of these possibilities can be constructed in one go in a ‘one-pot reaction’. When different combinations of regulatory parts are tried out to explore the expression space it is

called a ‘combinatorial approach’, and this has been used recently to optimise production of the therapeutic natural product violacein in *S. cerevisiae* [69].

2.2.2.2 Protein Engineering and Directed Evolution

Changing the expression levels or the spatial organisation of pathway enzymes can improve biosynthesis yields, but changing the amino acid sequences of the enzymes can make even more dramatic improvements in yield. For example by increasing the specificity of an enzyme for its intended substrate, unwanted catalysis of wasteful side reactions can be reduced, improving biosynthesis yields. Enzyme improvements like this can be achieved by mutagenesis of the coding sequence of the enzyme-encoding gene, however, it is often difficult to do this in a completely rational way, largely because that requires detailed understanding of the enzyme 3D structure. The number of solved protein structures for enzymes is relatively low compared to known protein sequences [70], and even when structures are known, catalytic site prediction is not trivial [71]. This makes it difficult to identify the important amino acid residues that could be mutated.

Instead if a simple test exists for enzyme activity, then enzymes can be more easily optimised by screening many designed or random gene variants that lead to a different mutations within the enzyme. When this process of mutagenesis and screening is repeated several times, with the best variants from one round selected as starting points for the next round, this is known as directed evolution [72]. In the recent case of production of precursors for the benzylisoquinoline alkaloids (BIAs), an elegant screen was used to optimise one of the pathway enzymes for dopamine production in the host organism *S. cerevisiae*. A promiscuous pathway enzyme that catalysed the desired conversion of L-tyrosine to L-DOPA, also catalysed an undesired reaction siphoning L-DOPA away to L-Dopaquinone. To address this problem, DeLoache and co-workers developed a biosensor-based screen utilising another enzyme (DOD) that convert L-DOPA into the yellow fluorescent pigment, betaxanthin [73]. By mutating their pathway enzyme in the presence of the DOD enzyme, they could screen for improved L-DOPA production by looking for increased yellow pigment accumulation. A mutant library over 200,000 variants of the pathway enzyme generated by error-prone PCR was screened and the best mutants selected for another round of mutagenesis and screening. This resulted in the identification of two amino acid mutations that together resulted in a 4.3-fold increase pigment production. The final mutated enzyme (a tyrosine hydroxylase) was shown to increase the amounts of intracellular L-DOPA available for Dopamine biosynthesis 2.8-fold versus the original enzyme, thanks to mutations that lowered the activity of the detrimental side reaction.

In synthetic biology terms, something that detects the presence of a molecule and is able to produce an easily-detectable output in response is called a biosensor [74]. For DeLoache and co-workers, their biosensor (an enzyme in this case) already existed in nature and formed the basis of their screen. However, often a convenient biosensor is not readily available for the purposes of

directed evolution of a natural product pathway. To address this synthetic biologists can develop novel biosensors or identify proteins or nucleic acids from other organisms that can be repurposed to work in biosensors. Many proteins detect metabolites and these natural roles can be exploited. For example, *E. coli* has over 200 transcription factors and many of these can sense different metabolites. Sometimes these can be engineered to detect a desired metabolite too [75]. RNA-based biosensors are also a useful tool. A recent study used an RNA-based sensor called a ribozyme (RNA that self-cleaves when it binds to its ligand) in order to screen for better producers of N-acetyl glucosamine, a small molecule with therapeutic properties [76]. In this study the ribozyme detects levels of intracellular glucosamine 6-phosphate (GlcN6P), a precursor to N-acetyl glucosamine (GlcNAC). Placing this ribozyme in the 3' untranslated region of an essential gene in the host organism, *S. cerevisiae*, means that self-cleavage leads to cell death. This setup was used to screen a mutant library of variants of an enzyme called GFA1, which converts GlcN6P to GlcNAC. Any enzyme variants that did not efficiently convert GlcN6P to GlcNAC would leave high levels of GlcN6P free to bind to the ribozyme and trigger cleavage of the essential gene's mRNA. Coupling mutation with this screen resulted in selecting a variant of GFA1 five times more efficient than its original version.

2.2.2.3 Genome Engineering

A complementary approach to optimising the output from a natural product pathway is to optimise the host itself. The least rational way of doing this is to use random mutagenesis coupled with screening for the desired property. This approach was used to optimise fungal strains for the manufacture of one of the first commercially produced natural products – penicillin [11]. This forward mutagenesis is completely random in terms of what genes will be mutated in the host genome and contrasts to targeted mutagenesis described in the sections above, where a single gene or region with this is altered. More recently, synthetic biology methods have been developed that allow much more directed genome mutagenesis. One of these methods, popular in *E. coli*, is called multiplexed automated genome engineering (MAGE) [77] and relies on hijacking the cell's DNA replication machinery. Large number of short pieces of single-stranded DNA (oligos) with a few mismatches to the genome sequence are added to the cell and allowed to incorporate into daughter chromosomes during DNA replication. The sequences of these oligos determine where they make base pair changes in the genome, and allow for simultaneous mutagenesis of multiple loci. By varying the number of oligos and their base sequences, researchers can mutate both the number of genomic sites to be mutated and the degree of saturation of mutation. Combined with genome-scale metabolic models [78, 79] and computational tools to predict which host genomic changes will be most effective for aiding greater production of a heterologous genetic pathway [80-82], genome-scale engineering using MAGE can be a very powerful tool. Recently this approach was used to effectively increase the production of resveratrol, a natural product of

therapeutic value, by making genome-wide mutations predicted to increase production of the pathway precursor malonyl-CoA [83].

Similar to MAGE, the more recently developed CRISPR-Cas9 technology for directed and multiplex chromosome cutting also allows multiple simultaneous genetic changes to be made to genomes [84, 85]. Unlike MAGE, the CRISPR-Cas9 system is universal to many organisms amenable to genetic transformation, opening up the use of genome engineering to optimise many hosts for natural product pathways. Typically, the CRIPSR-Cas9 system consists of two components: a protein and an RNA. The protein component is the Cas9 endonuclease that cuts DNA. The RNA component, termed the guide RNA, directs Cas9 to a genomic locus by Watson-Crick base pairing. This latter feature makes the system far easier to target to particular genomic loci than previous genome editing endonuclease technologies such as TAL-effectors and zinc finger nucleases [86].

The Cas9 system can also be modified to not cut DNA (CRISPR-dCas9). When fused to transcriptional activator or repressor protein domains, dCas9 can target genomic promoters to activate or repress gene expression. .This potentially enables multiplex and targeted alteration of host gene expression, which could be used to optimise the host for pathway output. Indeed, this approach has recently been applied to demonstrate fine control over pathway intermediates in the production of the therapeutic natural product violacein [87]. In this important work Zalatan and co-workers altered the paradigm above to make it easier to engineer. Instead of fusing transcriptional activator or repressor domains to the dCas9 protein, transcriptional modulating capabilities were engineered into the guide RNA component of the CRISPR-dCas9 system. This was achieved by extending the guide RNAs with scaffold RNAs that acted as docking sites for RNA-binding protein-transcriptional modulator protein fusions. Using these augmented guide RNAs and fusion proteins in concert with dCas9 expressed under a galactose-inducible promoter, Zalatan and colleagues were able to easily induce entire transcriptional programs in engineered *S. cerevisiae* cells. In this manner, complete control was achieved over the violacein biosynthetic pathway in the heterologous host, with high production of each of the possible pathway intermediates depending on which combination of augmented guide RNAs was expressed.

The field of genome editing with CRISPR is rapidly evolving, and new developments include the use of alternate endonucleases such as Cpf1, that allow greater convenience and specificity in certain cases [88]. For more in-depth discussion of CRISPR in a metabolic engineering context, the reader is pointed to the following recent review [89].

3. Unknown and Unnatural Natural Products

The examples discussed above use synthetic biology to improve the production of therapeutic natural products with known biosynthetic pathways. The following sections address cases where

synthetic biology is used to discover new (unknown products) therapeutic natural products in nature, or to create therapeutic molecules not found in nature (unnatural products).

3.1 Unknown Natural Products

The biosynthetic genes that encode each of the major classes of natural products illustrated in Figure 1 often have signature features that are common to most members of the class. Biosynthesis of the terpenoids, alkaloids, nonribosomal peptides and polyketides involves the production of a chemical scaffold molecule, which is often decorated by the addition of functional groups, or by cyclisation and branching to create a plethora of possible molecules [90-92]. Each of these four major classes of natural product has signature biosynthetic genes encoding enzymes specific to that class, and often these genes physically cluster together in genomes, as illustrated in Figure 3a [93, 94].

These two features allow the computational prediction of natural product biosynthesis genes in sequenced genomes. Indeed, probabilistic computational tools and standards have been developed to automate this prediction in genome sequences, even when the genome is otherwise unannotated [95, 96]. Given that most of our therapeutic drugs are derived from natural products, the discovery of novel natural products is a promising avenue for the development of novel therapeutic drugs. The situation where natural product genes can be predicted means that currently “unknown natural products” can be inferred, and their chemical structure predicted before the molecule is ever detected.

3.1.1 Genomics

With the advent of the genomic era and abundant cheap DNA sequencing, it was realised that the genomes of some of the most prolific producers of therapeutic natural products, such as the *Streptomyces* bacteria and *Aspergillus* fungi, contained many more biosynthetic gene clusters than the number of natural products known to be made by these species [97]. One explanation for this discrepancy is that in the native organism natural products are often associated with specific developmental stages or environmental conditions, and that these stages or conditions are not elicited under laboratory growth conditions [98].

To this end, synthetic biology strategies have been applied to ‘awaken’ these silent natural product gene clusters. These strategies have relied on adding or removing genetic control elements (e.g. promoters) to the cluster, in an effort to bypass any native regulation that keeps the cluster silent. Figure 3b depicts this process, which is known as ‘refactoring’ the cluster [9]. Refactoring can be done within the native natural product host organism if it is amenable to genetic manipulation, or by cloning the entire cluster, with modifications, to a heterologous host. One type

of target for refactoring is cluster-specific genes that regulate transcription of the natural product genes. For transcriptional repressors, the idea is to delete these genes, and the opposite is true for transcriptional activators. A recent example demonstrating the utility of this approach concerned a nonribosomal peptide gene cluster found in a marine bacterium, the actinomycete *Saccharomonospora* sp. CNQ490. Computational analysis of the draft genome of this organism identified 19 putative natural product gene clusters, most of which were silent clusters, and only one of which only one had an associated known product. Upon identifying a silent cluster predicted to produce a nonribosomal peptide, Yamanaka and co-workers set out to refactor the cluster by removing a negative regulatory gene [99]. Using a "transformation and recombination cloning" (TAR cloning), the authors deleted a cluster gene predicted to encode a repressive transcription factor. The resulting mutant cluster (but not the wild type cluster) produced a series of novel chlorinated lipopeptides similar to the known antibiotic daptomycin.

A related approach involves modifying the promoters of cluster-specific biosynthetic genes, rather than transcription factors residing within clusters. This approach was recently used in a study that added arabinose-inducible promoters to allow the cluster to be switched on as desired. This refactoring enabled activation of a silent gene cluster from the symbiotic bacterium *Xenorhabdus budapestensis* that produced a compound similar to the antibiotic holomycin in the [100]. Examples of this type demonstrate the utility of awakening silent clusters towards discovering novel therapeutic molecules.

Another trend in the genomics of 'unknown' therapeutic natural products has been to sequence the genomes and transcriptomes of plants used in traditional medicine. The motivation in these cases is that these organisms are more likely to be enriched for novel therapeutic natural products. While there are several recent studies using this approach to identify candidate biosynthetic genes for therapeutic alkaloids and terpenoids [92, 101, 102], synthetic biology approaches to produce these compounds have lagged behind. One potential reason for this situation is the apparent observation that plant therapeutic natural product genes are not as often physically clustered in the genome as their counterparts in bacteria and fungi [103], making their computational identification more difficult.

3.1.2 Metagenomics

In addition to sequencing the genomes of individual organisms, the falling costs of DNA sequencing and improving computational methods mean that whole communities of microorganisms in environmental soil, in the human gut and in ocean samples can now be sequenced as one [9, 93]. Analogously to the situation for individual genomes, silent natural product pathways can be identified from these metagenomes. In this case it is unclear that these pathways are silent in their hosts since it is not clear what each organism is within such a mixed environmental sample: typically these are organisms that have never been grown in lab conditions.

These clusters are appropriately called ‘cryptic’ clusters and refactoring approaches have successfully enabled expression of these pathways in heterologous hosts. For example, Brady and co-workers mined metagenomes of diverse microbial communities associated with extreme conditions (i.e. drought, high temperatures and salinity) and found and heterologously expressed novel polyketide and nonribosomal peptide gene clusters and successfully produced new antibiotic, and anti-tumour agents and an immunosuppressant [104, 105]. As more metagenomes become available, this approach will likely become increasingly important.

3.2. Unnatural Products

A long-term goal of natural product research has been to create new ‘unnatural’ variants of natural products by understanding and manipulating the biosynthetic logic behind certain classes of natural product. It has been reasoned that tweaking the biosynthetic enzymatic machinery appropriately would create novelty while retaining the chemical properties that make natural products valuable as therapeutic drugs.

3.2.1 Incorporating Unnatural Precursors

The biosynthesis of natural product genes of all four of the main classes shown in Figure 1 begins with incorporation of simple chemical building blocks, or monomers into a larger chemical backbone. One strategy for generating unnatural natural products has been to feed the host organisms with unnatural, synthetic variants of the natural building blocks, or precursors, as can be seen in Figure 4a. This approach relies on the ability of the natural product biosynthetic enzymes to tolerate non-cognate substrates. This approach was recently used to explore the substrate promiscuity of a polyketide biosynthetic enzyme that normally produces rhizoxin, a potent antimitotic agent [106]. By expressing the polyketide biosynthetic gene in a heterologous host and feeding different synthetic monomers, they were able to change the final ‘unnatural product’ into molecules that more closely resembled the antibiotic cyclohexamide. Similarly, pioneering work from the O’ Connor group has extended this paradigm to alkaloid natural products. The Madagascar periwinkle (*C. roseus*) is a medicinal plant that produces over 100 different alkaloids, including the anti-cancer agents vincristine and vinblastine [92]. By feeding this plant synthetic, derivatised versions of tryptamine, the natural building block for terpene indole alkaloids (TIAs), McCoy and co-workers were able to generate potentially bioactive unnatural TIAs [107].

This approach can also be used for developing completely new drugs. Asai and co-workers recently discovered a novel polyketide biosynthetic gene in the fungus *Chaetomium*, and expressed this pathway in the heterologous host *A. oryzae*. This pathway was found to produce several polyketides via the production of a very reactive intermediate which was exploited by feeding the heterologous host a library of chemically synthesised molecules, resulting in a set of

structurally diverse ‘pseudo-natural products’ [108]. Importantly, one of these products showed potent activity against adenovirus, a highly infective virus family with few current treatments. This demonstrates the utility of harnessing natural product biosynthetic machinery to produce new molecules that potentially can become new drugs.

A complementary strategy has been to mutate natural product biosynthetic enzymes to increase their tolerance for non-cognate substrates. A single mutation in a nonribosomal peptide biosynthetic gene has been shown to relax its substrate specificity and allow incorporation of non-natural amino acids functionalised with azide and alkyne groups [109]. These groups enable the final product molecules to be used in CLICK chemistry, which can be used to further diversify unnatural products to improve their therapeutic value by increasing bioactivity or reducing toxicity [110].

3.2.2 Combinatorial Biosynthesis

In the context of natural products, ‘combinatorial biosynthesis’ is the bringing together of enzymes from different biosynthetic pathways to produce molecules not found in nature. An overview of this approach is outlined in Figure 4c. Synthetic biology is particularly well suited to this task as it involves either expression of enzymes in heterologous hosts or modification of enzymes in natural hosts. One example concerns saponins, a type of terpenoid. Saponins are structurally diverse bioactive compounds composed of 30-carbon scaffolds that are decorated with multiple functional groups and sugars. One of the key functional groups is the hydroxyl group, added by a family of enzymes called cytochromes p450. In a recent study, Moses and co-workers cloned a novel cytochrome p450 from a plant used in Asian traditional medicine and used synthetic biology approaches to express it along with the saponin scaffold biosynthetic enzymes and glycosyltransferases in the heterologous host *S. cerevisiae* [111]. In this manner they were able to generate unnatural, potentially bioactive saponins decorated with new combinations of hydroxyl and sugar groups.

In natural product biosynthetic pathways in general, during and after the production of the main chemical backbone molecule via the linking of monomers, further chemical diversity is often achieved by the action of specialised tailoring enzymes [112]. These enzymes introduce small chemical modifications, such as the addition of methyl, hydroxyl, sulfyl and glycosyl groups. By expressing independent tailoring enzymes in conjunction with the main biosynthetic genes, potentially unnatural diversification of the backbone molecules can be achieved. These modifications are often important for conferring bioactivity or for removing toxicity to human cells [113]. In a recent study by Yin and co-workers, a class of tailoring enzymes was discovered by computational analysis of the draft genome of an *Actinoplanes* fungus. The authors used two of these enzymes to add sulphate and glycosyl groups to a nonribosomal peptide antibiotic produced by a different organism, the bacterium *Streptomyces toyocaensis* [114]. Another study expanded

this paradigm to include tailoring reactions not normally present in natural products. In this seminal work, Walker and co-workers exploited enzymes in the only biological pathway known to utilise fluorine to site-specifically incorporate fluorine into two polyketides *in vivo* [115]. In a similar vein, Runguphan and co-workers expressed chlorination enzymes from soil bacteria in the Madagascar periwinkle, a medicinal plant with prolific alkaloid production capabilities, to yield chlorinated alkaloids [116]. By using enzymes to incorporate chemical changes once exclusively in the realm of synthetic chemistry (fluorination and chlorination), these studies represent an important step forward in the application of synthetic biology to expand the diversity of natural products.

For two of the major classes of natural products, the polyketides and the nonribosomal peptides, the biosynthetic enzymes responsible for the synthesis of the natural product backbone are modular. Further, the number and order of modules within the gene usually corresponding to the number and order of chemical building blocks in the backbone molecule that becomes the natural product [112, 117]. For these natural products, instead of combining different biosynthetic enzymes as described above, combinatorial biosynthesis can be achieved by swapping modules within a single enzyme as shown in Figure 4b. Thus protein engineering has been used to swap modules in polyketide and nonribosomal peptide biosynthetic genes in further attempts to generate unnatural products. This approach was used to make variants of the antibiotic daptomycin by exploring module exchanges in dptD, a gene encoding a di-modular enzyme that incorporates the monomers 3-methylglutamic acid and kynurenine into daptomycin. By replacing the kynurenine-specific module with modules specifying asparagine, the engineered nonribosomal peptide synthetase now produces a structural variant of daptomycin with similar antibiotic properties [118]. The ability to confer such structural diversity to existing antibiotics may prove fruitful in the fight against antibiotic resistance.

For the most part, combinatorial biosynthesis for therapeutic drug development has focussed on a single class of natural product. Recently however, an ambitious application of synthetic biology for production of natural-product derived therapeutic molecules combined biosynthetic enzymes from several different natural product classes. Klein and co-workers cloned a myriad of biosynthetic enzymes from alkaloid, polyketide and flavonoid (a minor natural product class) pathways, and expressed dozens of different combinations of these in the yeast *S. cerevisiae*, using a novel recombination-based approach [119]. Over 75% of the resulting 74 compounds were new to science, and 20% were highly different, representing novel chemical backbones. All of these exhibited structural complexity akin to that of natural products, and were enriched for several other metrics of drug-likeness.

4. Future Outlook

One key area of synthetic biology research concerns designing ‘synthetic cells’. One can envision a future where synthetic cells are rationally tailored for the production of small molecules, with

customised cells made to fit the biosynthetic pathway in question. Efforts are already being made in this direction, as evidenced by the recent construction of 'modular cells' and their application for the production of various chemicals in customised *E. coli* [120]. Aside from synthetic genomes, the synthetic cell concept also applies to subcellular compartments. Already progress has been made to recapitulate naturally occurring bacterial micro-compartments (BMCs) and viral capsids into the model host bacterium *E. coli* [121]. While both of these types of natural compartment have been used to aid the production of small molecules the next conceptual step would be to design synthetic sub-cellular compartments from the ground up, tailored to the optimal reaction conditions of specific natural product pathways.

Another avenue being explored by synthetic biology takes an opposite conceptual direction. Instead of adding compartments, in cell-free synthetic biology the aim is to remove the principle compartment and carry out biological processes with defined cellular components. Cell-free systems offer the advantages of building biological systems from the ground up for applications such as metabolite synthesis rather than adapting existing biological systems for the task. One major advantage of such an approach is that there is no need to optimise the background genome of a host organism for better production. Further, it is often easier to fine-tune the performance of multi-enzyme pathways such as those for most natural products in cell-free systems, as systems-level properties can be directly monitored in real-time [122].

For unknown natural products, future efforts will likely bring better computational methods for characterising biosynthetic clusters from genomic and metagenomic DNA sequences. Studies of the diversity of natural products in geographically distinct soil samples from around the world suggest that natural product diversity is potentially much greater than appreciated from microorganisms amenable to laboratory based culture [93]. Advances in molecular biology methods for cloning and expressing these biosynthetic pathways should open up new routes to harnessing the potential of natural products from the "microbial dark matter". In particular, methods for cheaper, more efficient de novo DNA synthesis will greatly bolster efforts to express biosynthetic pathways in heterologous hosts. For a more detailed discussion about advances in this area, readers are pointed to the following recent review [123]. Finally, for "unnatural products", progress is being made towards understanding the rules for rational redesign of the biosynthetic machinery towards novel molecules with drug-like properties. Synthetic biology is likely to play a key role in making enzyme design predictable.

5. Conclusions

Natural products have provided the basis for many of the most important drugs developed in the past century. However, the last two decades have witnessed a decrease in natural product based drug development by the pharmaceutical industry in favour of synthetic small molecule library screening. Nonetheless, the application of synthetic biology approaches such as those discussed

738 in this review holds great promise for reviving and enhancing this great natural resource for our
739 benefit in coming centuries, and also going beyond nature to produce new therapeutic molecules
740 too.
741
742
743
744 **Acknowledgments**
745 Research on therapeutic natural product biosynthesis by A.R.A and T.E. is supported by the UK
746 Biotechnology and Biological Sciences Research Council (BBSRC) grant BB/K006290/1. W.S. is
747 supported by a BBSRC AstraZeneca CASE PhD studentship award (BB/M503381/1). The authors
748 would like to thank Dr. Ciarán Kelly for very helpful feedback about the manuscript style and
749 content.
750

References

[1] H. Lévesque, O. Lafont, *Rev Med Interne* (2000) **21 Suppl 1**, 8s-17s.

[2] W. Mitchell, *Curr Opin Chem Biol* (2011) **15**, 505-15.

[3] D. J. Newman, G. M. Cragg, *J Nat Prod* (2007) **70**, 461-77.

[4] D. J. Newman, G. M. Cragg, *J Nat Prod* (2012) **75**, 311-35.

[5] A. L. Demain, R. P. Elander, *Antonie Van Leeuwenhoek* (1999) **75**, 5-19.

[6] M. S. Butler, A. D. Buss, *Biochem Pharmacol* (2006) **71**, 919-29.

[7] F. Peláez, *Biochem Pharmacol* (2006) **71**, 981-90.

[8] E. Andrianantoandro, S. Basu, D. K. Karig, R. Weiss, *Mol Syst Biol* (2006) **2**, 2006.0028.

[9] H. -J. Frasch, M. H. Medema, E. Takano, R. Breitling, *Curr Opin Biotechnol* (2013) **24**, 1144-50.

[10] J. D. Keasling, *Metab Eng* (2012) **14**, 189-95.

[11] L. B. Pickens, Y. Tang, Y. -H. Chooi, *Annu Rev Chem Biomol Eng* (2011) **2**, 211-36.

[12] A. J. Hockenberry, M. C. Jewett, *Curr Opin Chem Biol* (2012) **16**, 253-9.

[13] Y. Li, B. A. Pfeifer, *Curr Opin Plant Biol* (2014) **19**, 8-13.

[14] A. Krivoruchko, V. Siewers, J. Nielsen, *Biotechnol J* (2011) **6**, 262-76.

[15] D. -K. Ro, E. M. Paradise, M. Ouellet, K. J. Fisher, *et al.*, *Nature* (2006) **440**, 940-3.

[16] M. H. Medema, R. Breitling, E. Takano, *Methods Enzymol* (2011) **497**, 485-502.

[17] V. Meyer, B. Wu, A. F. J. Ram, *Biotechnol Lett* (2011) **33**, 469-76.

[18] Y. Zhuo, T. Zhang, Q. Wang, P. Cruz-Morales, *et al.*, *Biotechnol J* (2014) **9**, 316-25.

[19] Y. Luo, L. Zhang, K. W. Barton, H. Zhao, *ACS Synth Biol* (2015) **4**, 1001-10.

[20] C. Bai, Y. Zhang, X. Zhao, Y. Hu, *et al.*, *Proc Natl Acad Sci U S A* (2015) **112**, 12181-6.

[21] Y. Luo, B. -Z. Li, D. Liu, L. Zhang, *et al.*, *Chem Soc Rev* (2015) **44**, 5265-90.

[22] M. R. Seyedsayamdost, J. Clardy, *ACS synthetic biology* (2014) **3**, 745-7.

[23] J. D. Keasling, *Science* (2010) **330**, 1355-8.

[24] J. -F. Martín, R. V. Ullán, C. García-Estrada, *J Ind Microbiol Biotechnol* (2012) **39**, 367-82.

[25] L. Gidijala, I. J. van der Klei, M. Veenhuis, J. A. K. W. Kiel, *FEMS Yeast Res* (2007) **7**, 1160-7.

[26] L. Gidijala, J. A. K. W. Kiel, R. D. Douma, R. M. Seifar, *et al.*, *PLoS One* (2009) **4**, e8317.

[27] C. -H. Yun, S. -K. Yim, D. -H. Kim, T. Ahn, *Current drug metabolism* (2006) **7**, 411-29.

[28] E. Leonard, M. A. Koffas, *Applied and environmental microbiology* (2007) **73**, 7246-51.

[29] I. Pateraki, A. M. Heskes, B. Hamberger, *Adv Biochem Eng Biotechnol* (2015) **148**, 107-39.

[30] Z. Dai, B. Wang, Y. Liu, M. Shi, *et al.*, *Sci Rep* (2014) **4**, 3698.

[31] K. Zhou, K. Qiao, S. Edgar, G. Stephanopoulos, *Nat Biotechnol* (2015) .

[32] F. Ceroni, R. Algar, G. -B. Stan, T. Ellis, *Nat Methods* (2015) **12**, 415-8.

786 33. D. A. Hopwood, *Natural product biosynthesis by microorganisms and plants* (Academic
787 Press, 2012).
788 [34] F. Y. Lim, N. P. Keller, *Nat Prod Rep* (2014) **31**, 1277-86.
789 [35] T. Y. Kim, S. B. Sohn, Y. B. Kim, W. J. Kim, S. Y. Lee, *Curr Opin Biotechnol* (2012) **23**, 617-23.
790 [36] Z. Dai, J. Nielsen, *Curr Opin Biotechnol* (2015) **36**, 8-15.
791 [37] J. Almquist, M. Cvijovic, V. Hatzimanikatis, J. Nielsen, M. Jirstrand, *Metab Eng* (2014) **24**,
792 38-60.
793 [38] J. W. Lee, D. Na, J. M. Park, J. Lee, *et al.*, *Nat Chem Biol* (2012) **8**, 536-46.
794 [39] K. A. Oye, J. C. H. Lawson, T. Bubela, *Nature* (2015) **521**, 281-3.
795 [40] I. J. Trenchard, M. S. Siddiqui, K. Thodey, C. D. Smolke, *Metab Eng* (2015) **31**, 74-83.
796 [41] J. E. Dueber, G. C. Wu, G. R. Malmirchegini, T. S. Moon, *et al.*, *Nat Biotechnol* (2009) **27**,
797 753-9.
798 [42] X. Chen, P. Zhu, L. Liu, *Metab Eng* (2015) .
799 [43] G. Sachdeva, A. Garg, D. Godding, J. C. Way, P. A. Silver, *Nucleic Acids Res* (2014) **42**, 9493-
800 503.
801 [44] M. H. Saier, *J Mol Microbiol Biotechnol* (2013) **23**, 243-69.
802 [45] T. W. Giessen, P. A. Silver, *J Mol Biol* (2015) .
803 [46] U. Heinig, M. Gutensohn, N. Dudareva, A. Aharoni, *Curr Opin Biotechnol* (2013) **24**, 239-
804 46.
805 [47] P. D. Straight, M. A. Fischbach, C. T. Walsh, D. Z. Rudner, R. Kolter, *Proc Natl Acad Sci U S A*
806 (2007) **104**, 305-10.
807 [48] M. Farhi, E. Marhevka, T. Masci, E. Marcos, *et al.*, *Metab Eng* (2011) **13**, 474-81.
808 [49] J. A. N. Brophy, C. A. Voigt, *Nature Methods* (2014) **11**, 508-20.
809 [50] J. Monod, J. -P. Changeux, F. Jacob, *Journal of molecular biology* (1963) **6**, 306-29.
810 [51] B. A. Blount, T. Weenink, S. Vasylechko, T. Ellis, *PLoS One* (2012) **7**, e33279.
811 [52] T. Ellis, X. Wang, J. J. Collins, *Nat Biotechnol* (2009) **27**, 465-71.
812 [53] M. H. Medema, R. Breitling, R. Bovenberg, E. Takano, *Nat Rev Microbiol* (2011) **9**, 131-7.
813 [54] E. Takano, J. White, C. J. Thompson, M. J. Bibb, *Gene* (1995) **166**, 133-7.
814 [55] W. Wang, X. Li, J. Wang, S. Xiang, *et al.*, *Appl Environ Microbiol* (2013) **79**, 4484-92.
815 [56] A. Rodríguez-García, P. Combes, R. Pérez-Redondo, M. C. A. Smith, M. C. M. Smith, *Nucleic*
816 *Acids Res* (2005) **33**, e87.
817 [57] E. K. Zess, M. B. Begemann, B. F. Pfleger, *Biotechnol Bioeng* (2015) .
818 [58] S. Kim, K. Lee, S. -J. Bae, J. -S. Hahn, *Appl Microbiol Biotechnol* (2015) **99**, 2705-14.

[59] C. J. Robinson, H. A. Vincent, M. -C. Wu, P. T. Lowe, *et al.*, *Journal of the American Chemical Society* (2014) **136**, 10615-24.
[60] J. J. Trausch, R. T. Batey, *Methods Enzymol* (2015) **550**, 41-71.
[61] R. Penchovsky, *Biomacromolecules* (2013) **14**, 1240-9.
[62] A. Espah Borujeni, D. M. Mishler, J. Wang, W. Huso, H. M. Salis, *Nucleic Acids Res* (2016) **44**, 1-13.
[63] A. Chakrabarty, G. T. Buzzard, A. E. Rundell, *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* (2013) **5**, 181-203.
[64] R. Kitney, P. Freemont, *FEBS letters* (2012) **586**, 2029-36.
[65] C. Engler, R. Kandzia, S. Marillonnet, *PLoS One* (2008) **3**, e3647.
[66] A. Casini, M. Storch, G. S. Baldwin, T. Ellis, *Nat Rev Mol Cell Biol* (2015) **16**, 568-76.
[67] E. Weber, C. Engler, R. Gruetzner, S. Werner, S. Marillonnet, *PLoS One* (2011) **6**, e16765.
[68] M. E. Lee, W. C. DeLoache, B. Cervantes, J. E. Dueber, *ACS Synth Biol* (2015) .
[69] M. E. Lee, A. Aswani, A. S. Han, C. J. Tomlin, J. E. Dueber, *Nucleic Acids Res* (2013) **41**, 10668-78.
[70] D. S. Marks, T. A. Hopf, C. Sander, *Nature biotechnology* (2012) **30**, 1072-80.
[71] J. E. Fajardo, A. Fiser, *BMC bioinformatics* (2013) **14**, 63.
[72] M. S. Packer, D. R. Liu, *Nature Reviews Genetics* (2015) .
[73] W. C. DeLoache, Z. N. Russ, L. Narcross, A. M. Gonzales, *et al.*, *Nat Chem Biol* (2015) **11**, 465-71.
[74] J. K. Rogers, C. D. Guzman, N. D. Taylor, S. Raman, *et al.*, *Nucleic Acids Res* (2015) **43**, 7648-60.
[75] D. Liu, T. Evans, F. Zhang, *Metab Eng* (2015) **31**, 35-43.
[76] S. -W. Lee, M. -K. Oh, *Metab Eng* (2015) **28**, 143-50.
[77] H. H. Wang, F. J. Isaacs, P. A. Carr, Z. Z. Sun, *et al.*, *Nature* (2009) **460**, 894-8.
[78] P. A. Jensen, K. A. Lutz, J. A. Papin, *BMC Syst Biol* (2011) **5**, 147.
[79] I. Nookaew, R. Olivares-Hernández, S. Bhumiratana, J. Nielsen, *Methods Mol Biol* (2011) **759**, 445-63.
[80] I. Rocha, P. Maia, M. Rocha, E. C. Ferreira, *OptGene: a framework for in silico metabolic engineering* (2008) .
[81] A. P. Burgard, P. Pharkya, C. D. Maranas, *Biotechnol Bioeng* (2003) **84**, 647-57.
[82] I. Rocha, P. Maia, P. Evangelista, P. Vilaça, *et al.*, *BMC Syst Biol* (2010) **4**, 45.
[83] N. Bhan, P. Xu, O. Khalidi, M. A. Koffas, *Chemical Engineering Science* (2013) **103**, 109-14.
[84] Y. Li, Z. Lin, C. Huang, Y. Zhang, *et al.*, *Metab Eng* (2015) **31**, 13-21.

[85] T. Jakočiūnas, I. Bonde, M. Herrgård, S. J. Harrison, *et al.*, *Metab Eng* (2015) **28**, 213-22.
[86] P. Mali, K. M. Esvelt, G. M. Church, *Nat Methods* (2013) **10**, 957-63.
[87] J. G. Zalatan, M. E. Lee, R. Almeida, L. A. Gilbert, *et al.*, *Cell* (2015) **160**, 339-50.
[88] B. Zetsche, J. S. Gootenberg, O. O. Abudayyeh, I. M. Slaymaker, *et al.*, *Cell* (2015) **163**, 759-71.
[89] T. Jakočiūnas, M. K. Jensen, J. D. Keasling, *Metab Eng* (2016) **34**, 44-59.
[90] A. A. Brakhage, *Nat Rev Microbiol* (2013) **11**, 21-32.
[91] F. K. Davies, R. E. Jinkerson, M. C. Posewitz, *Photosynth Res* (2015) **123**, 265-84.
[92] T. Dugé de Bernonville, M. Clastre, S. Besseau, A. Oudin, *et al.*, *Phytochemistry* (2015) **113**, 9-23.
[93] B. V. B. Reddy, D. Kallifidas, J. H. Kim, Z. Charlop-Powers, *et al.*, *Appl Environ Microbiol* (2012) **78**, 3744-52.
[94] J. M. Palmer, N. P. Keller, *Curr Opin Microbiol* (2010) **13**, 431-6.
[95] T. Weber, K. Blin, S. Duddela, D. Krug, *et al.*, *Nucleic Acids Res* (2015) **43**, W237-43.
[96] N. Khaldi, F. T. Seifuddin, G. Turner, D. Haft, *et al.*, *Fungal Genet Biol* (2010) **47**, 736-41.
[97] A. A. Brakhage, V. Schroeckh, *Fungal Genet Biol* (2011) **48**, 15-22.
[98] F. Y. Lim, J. F. Sanchez, C. C. C. Wang, N. P. Keller, *Methods Enzymol* (2012) **517**, 303-24.
[99] K. Yamanaka, K. A. Reynolds, R. D. Kersten, K. S. Ryan, *et al.*, *Proc Natl Acad Sci U S A* (2014) **111**, 1957-62.
[100] E. Bode, A. O. Brachmann, C. Kegler, R. Simsek, *et al.*, *Chembiochem* (2015) **16**, 1115-9.
[101] J. M. Hagel, J. S. Morris, E. -J. Lee, I. Desgagné-Penix, *et al.*, *BMC Plant Biol* (2015) **15**, 227.
[102] P. Daisy, S. S. Salu Mathew, N. A. Rayan, *International journal of biomedical science: IJBS* (2008) **4**, 196.
[103] B. Field, A. Osbourn, *Mob Genet Elements* (2012) **2**, 46-50.
[104] F. -Y. Chang, M. A. Ternei, P. Y. Calle, S. F. Brady, *J Am Chem Soc* (2015) **137**, 6044-52.
[105] J. G. Owen, B. V. B. Reddy, M. A. Ternei, Z. Charlop-Powers, *et al.*, *Proc Natl Acad Sci U S A* (2013) **110**, 11797-802.
[106] D. Heine, T. Bretschneider, S. Sundaram, C. Hertweck, *Angew Chem Int Ed Engl* (2014) **53**, 11645-9.
[107] E. McCoy, S. E. O'Connor, *J Am Chem Soc* (2006) **128**, 14276-7.
[108] T. Asai, K. Tsukada, S. Ise, N. Shirata, *et al.*, *Nature Chemistry* (2015) .
[109] H. Kries, R. Wachtel, A. Pabst, B. Wanner, *et al.*, *Angew Chem Int Ed Engl* (2014) **53**, 10105-8.
[110] N. Ma, Y. Wang, B. -X. Zhao, W. -C. Ye, S. Jiang, *Drug Des Devel Ther* (2015) **9**, 1585-99.

[111] T. Moses, J. Pollier, L. Almagro, D. Buyst, *et al.*, *Proc Natl Acad Sci U S A* (2014) **111**, 1634-9.
[112] C. T. Walsh, *Science* (2004) **303**, 1805-10.
[113] T. W. Giessen, M. A. Marahiel, *Front Microbiol* (2015) **6**, 785.
[114] G. Yim, L. Kalan, K. Koteva, M. N. Thaker, *et al.*, *Chembiochem* (2014) **15**, 2613-23.
[115] M. C. Walker, B. W. Thuronyi, L. K. Charkoudian, B. Lowry, *et al.*, *Science* (2013) **341**, 1089-94.
[116] W. Runguphan, X. Qu, S. E. O'Connor, *Nature* (2010) **468**, 461-4.
[117] H. D. Mootz, D. Schwarzer, M. A. Marahiel, *Chembiochem* (2002) **3**, 490-504.
[118] S. Doekel, M. -F. Coëffet-Le Gal, J. -Q. Gu, M. Chu, *et al.*, *Microbiology* (2008) **154**, 2872-80.
[119] J. Klein, J. R. Heal, W. D. Hamilton, T. Boussemghoune, *et al.*, *ACS Synthetic Biology* (2014) , 140116141820008.
[120] C. T. Trinh, Y. Liu, D. J. Conner, *Metab Eng* (2015) .
[121] F. Sargent, F. A. Davidson, C. L. Kelly, R. Binny, *et al.*, *Microbiology* (2013) **159**, 2427-36.
[122] C. E. Hodgman, M. C. Jewett, *Metab Eng* (2012) **14**, 261-9.
[123] S. Kosuri, G. M. Church, *Nat Methods* (2014) **11**, 499-507.

Figure Captions

Figure 1. The major classes of natural product. The four major classes of natural product are shown. For each class, a member with therapeutic properties is given along with the producing organism.

Figure 2. Optimising Pathway Yield. a) Rational and semi-rational strategies for optimising pathway yield in a heterologous host. Rational approaches include up-regulation of host metabolic genes that produce a pathway precursor, and down-regulating genes in pathways that compete for a synthetic pathway intermediate. Semi-rational approaches include using libraries of regulatory elements and pathway enzymes to explore gene expression and enzyme kinetic space. **b)** Spatial strategies include tethering pathway enzymes to a synthetic scaffold to control stoichiometry, and compartmentalisation of pathway enzymes to control reaction conditions and enzyme and substrate concentration. A third strategy involves a microbial co-culture of different species, each containing part of the pathway. Co-culture can be implemented in a mutualistic fashion to maintain both populations.

Figure 3. Mining (meta)genomic gene clusters for unknown natural products. a) It is increasingly simple to sequence the genomes of laboratory grown or environmentally sampled. These genomes sequences are searched for gene clusters indicative of unidentified natural products. **b)** These gene clusters are refactored for expression in a heterologous host to produce and identify the unknown natural product.

Figure 4. Creation of 'unnatural products'. a) Incorporating unnatural precursors. Substrate promiscuity of natural product biosynthetic enzymes can lead to the incorporation of unnatural precursors and chemical diversification of natural product backbones. **b)** Module swapping in polyketide and nonribosomal peptide biosynthetic machinery. The biosynthetic enzymes of two major classes of natural products are modular, with the number and order of modules specifying the identity of the natural product. By swapping, adding or deleting modules within these enzymes it is possible to create new compounds. **c)** Natural product biosynthetic genes from different organisms can be combined with 'tailoring enzymes' to produce structurally diverse 'unnatural products'.

Figure 1
Click here to download high resolution image

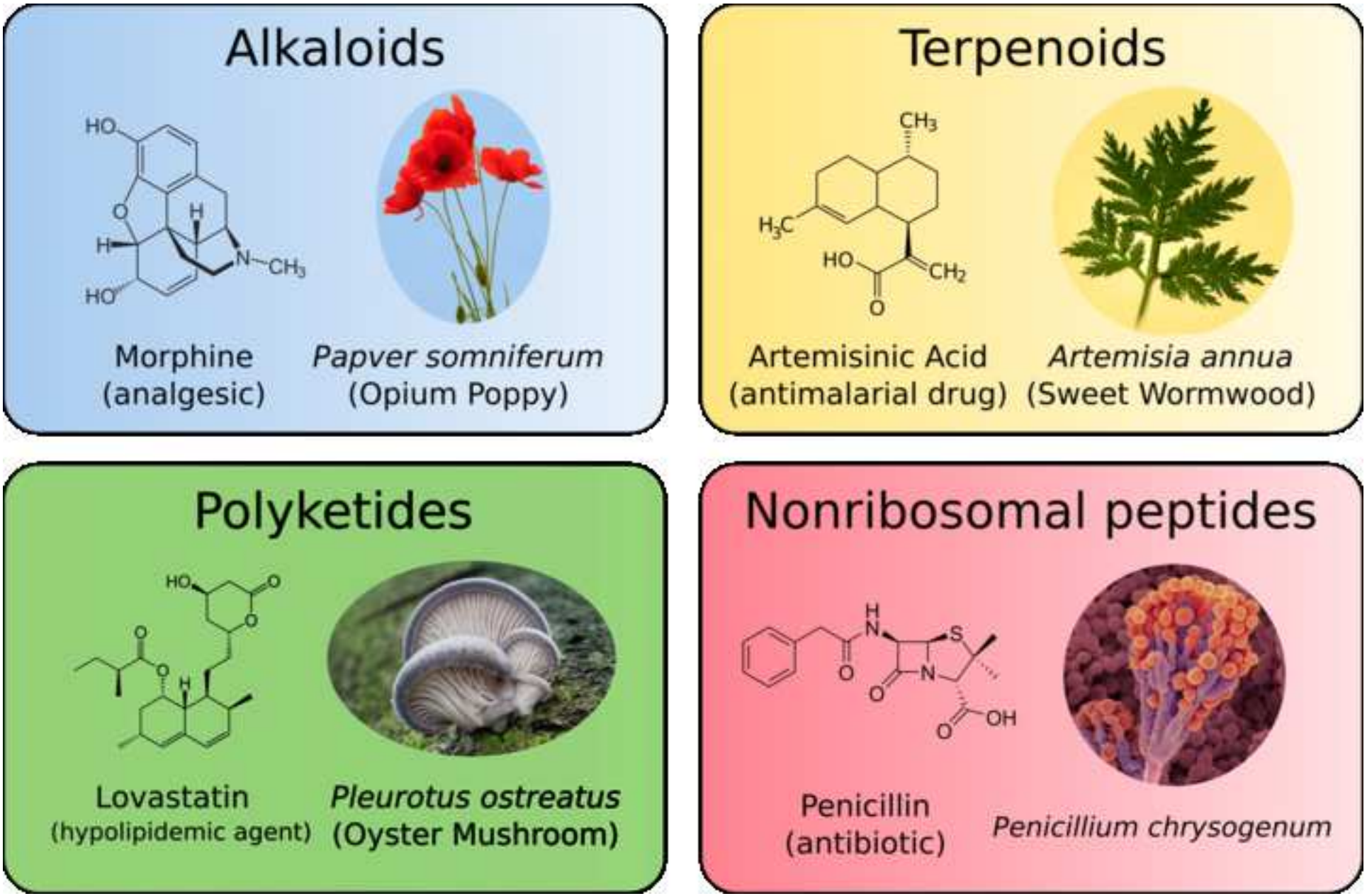

Figure 2

Click here to download high resolution image

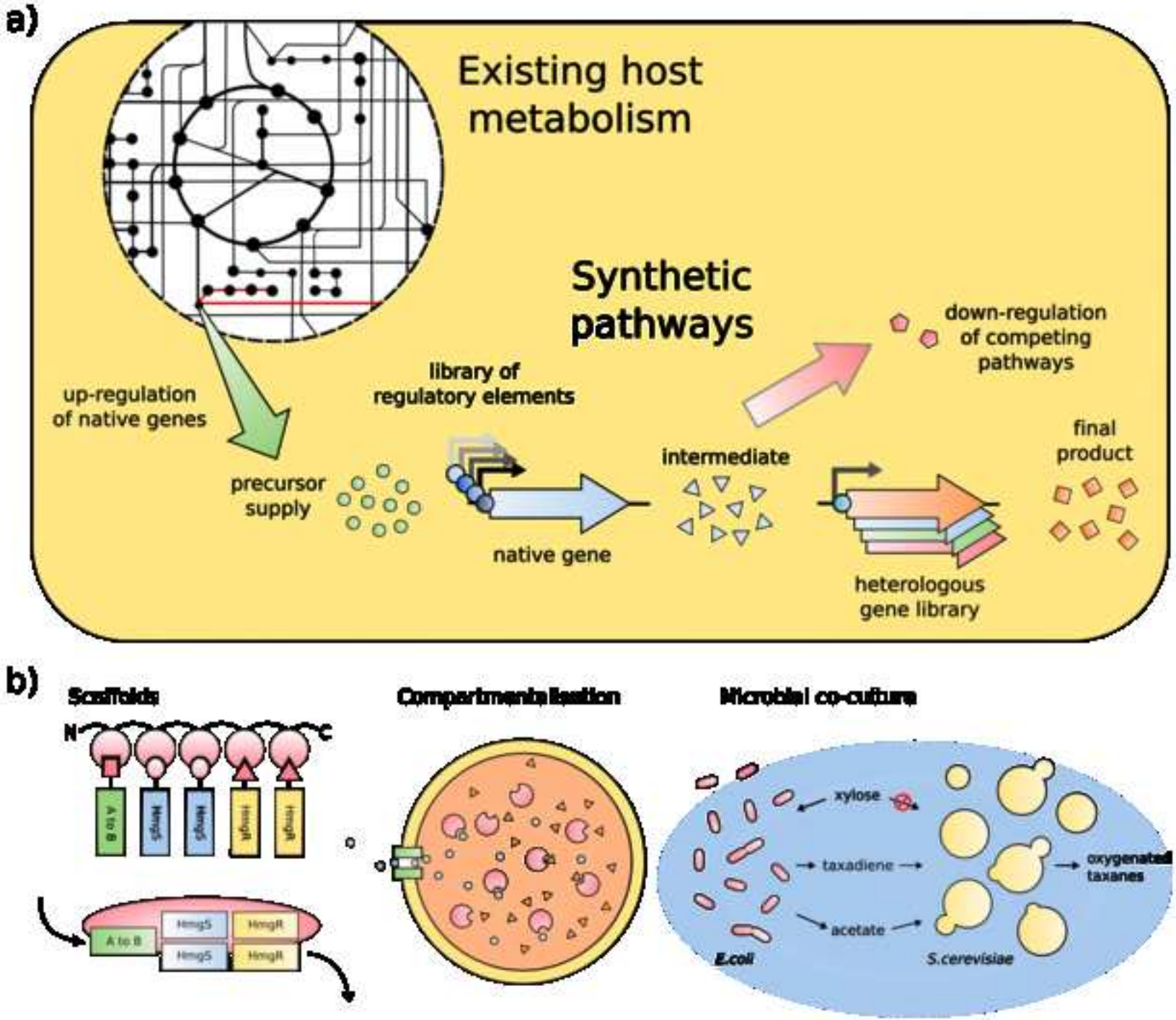

Figure 3

Click here to download high resolution image

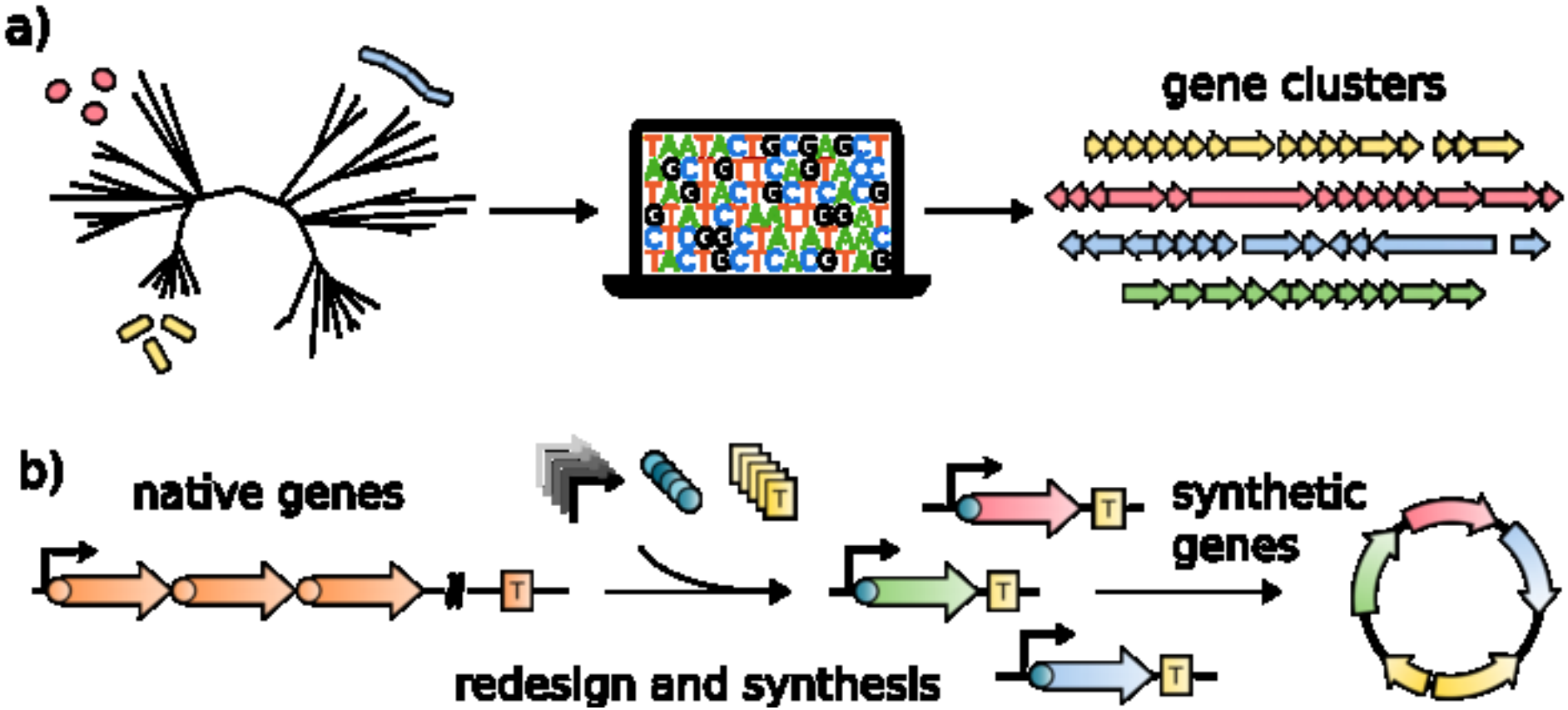

Figure 4

Click here to download high resolution image