J Ind Microbiol Biotechnol (2014) 41:469–477
DOI 10.1007/s10295-013-1358-3

REVIEW

SIMB
Society for Industrial Microbiology and Biotechnology

DNA assembly techniques for next-generation combinatorial biosynthesis of natural products

Ryan E. Cobb · Jonathan C. Ning · Huimin Zhao

Received: 14 August 2013 / Accepted: 25 September 2013 / Published online: 15 October 2013

Abstract Natural product scaffolds remain important leads for pharmaceutical development. However, transforming a natural product into a drug entity often requires derivatization to enhance the compound's therapeutic properties. A powerful method by which to perform this derivatization is combinatorial biosynthesis, the manipulation of the genes in the corresponding pathway to divert synthesis towards novel derivatives. While these manipulations have traditionally been carried out via restriction digestion/ligation-based cloning, the shortcomings of such techniques limit their throughput and thus the scope of corresponding combinatorial biosynthesis experiments. In the burgeoning field of synthetic biology, the demand for facile DNA assembly techniques has promoted the development of a host of novel DNA assembly strategies. Here we describe the advantages of these recently developed tools for rapid, efficient synthesis of large DNA constructs. We also discuss their potential to facilitate the simultaneous assembly of complete libraries of natural product biosynthetic pathways, ushering in the next generation of combinatorial biosynthesis.

R. E. Cobb · J. C. Ning · H. Zhao (✉)
Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
e-mail: zhao5@illinois.edu

R. E. Cobb · H. Zhao
Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

H. Zhao
Departments of Chemistry, Biochemistry, and Bioengineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

Keywords Combinatorial biosynthesis · DNA assembly · Genome mining · Synthetic biology

Introduction

In the modern era of pharmaceutical development and discovery, high-throughput screening methods are central to discovering new drugs. Some pharmaceutical companies have screening collections of sizes on the order of one million entities [5]. When screening such collections, hit rates can vary from 0.001 to 0.3 % [23], requiring many samples to be tested before a single hit is generated. Among the leads that become approved drugs, natural products or molecules derived from natural products represented 74.8 % of all cancer drugs approved by the FDA from 1981 to 2010 [33]. While evermore sensitive and accurate screening methods are continually being developed by academia and industry to increase the number of hits generated, there is also room for improvement in diversifying the screening libraries themselves. By and large, the natural product and extract libraries employed in modern high-throughput screens are still assembled using the same methods as in the 1970s. As such, many supposedly new hits often result in rediscovery of known compounds. One proposed method of expanding these libraries is combinatorial biosynthesis, whereby complex and novel products can be formed by taking a set of enzymes from disparate sources and combining their activities to form new biosynthetic pathways. This method is attractive due to the ability of biological catalysts to perform highly selective reactions that are difficult to achieve by chemical methods, giving it an advantage in product complexity and diversity over purely chemistry-based methods [47].

While in the past the prospects of combinatorial biosynthesis have been limited by the availability of known

enzymes to concatenate into pathways [43], in the current era of ever-decreasing genome sequencing costs there is a continually expanding selection of enzymes that can be utilized. Bioinformatics tools like the basic local alignment search tool (BLAST) [1] and protein domain recognition databases, including the protein families (Pfam) database and the conserved domains database (CDD), give rapid prediction of enzyme function without the need for laborious expression and isolation [28, 35]. Tools for the prediction of secondary metabolic pathways based on only genomic sequencing data also exist, which cover almost the entire scope of common secondary metabolites [9, 18, 30]. In addition, it is now realized that the number of secondary metabolite gene clusters present across all domains of life far exceeds the number of known secondary metabolites discovered under laboratory conditions. For example, although only a handful of natural products were previously known to be produced by *Streptomyces coelicolor* A3(2), *Streptomyces griseus* IFO 13350, and *Streptomyces avermitilis* MA-4680, genome sequencing revealed over 20 [3], 34 [34], and 38 secondary metabolite clusters [17], respectively, in these strains. The ability to rapidly identify a diverse set of homologous enzymes with potentially different catalytic activities or substrate specificities opens the door to the realization of combinatorial biosynthesis.

At the heart of most combinatorial biosynthesis efforts are the driving forces of potentially valuable and novel "non-natural natural products", knowledge of a set of enzymes amenable to combinatorial methods, and the ever-improving biological techniques to realize them. Up against these motivating factors are the challenges of natural protein–protein orthogonality, incompatible substrate scope, and limitations of microbial host capabilities and genetic manipulability. The story of combinatorial biosynthesis thus far, and very likely into the future, is that of the understanding of secondary metabolic pathway subtleties, and engineering efforts to create ever more robust expression systems and genetic tools. In this mini-review, we will highlight some of the major successes in combinatorial biosynthesis of natural products, and describe the host of new DNA assembly techniques that are poised to revolutionize the field.

Classic approaches to combinatorial biosynthesis

The initial attempts at combinatorial biosynthesis focused on the polyketide synthase (PKS) family of enzymes, due to the inherent modularity of PKS systems and the predictability of their end products [43]. Each PKS module catalyzes one specific step in polyketide synthesis before passing the maturing product onto the next module [43], making it straightforward to envision the combinatorial mixing and matching of these "assembly line"-type complexes to synthesize new non-natural products. Typically, one well-characterized system would be modified to create a small number of novel natural product derivatives. For instance, the erythromycin PKS system has been modified by the substitution of acyl-transferase domains of multiple PKSs to produce a library of 61 end products, many of which had not been observed in nature (Fig. 1) [29]. The pikromycin, tylosin and erythromycin PKS genes have also been combinatorially assembled to produce novel macrolides [44].

In addition to modifying enzymes in preexisting pathways, another combinatorial strategy is to combine enzymes to create de novo pathways in heterologous hosts. Two flavanones, pinocembrin and naringenin, were produced in *Escherichia coli* by expressing a phenylalanine ammonia-lyase from the fungus *Rhodotorula rubra*, a 4-coumarate:CoA ligase from the actinobacteria species *S. coelicolor*, the PKS chalcone synthase from the legume *Glycyrrhiza echinat*, and a chalcone isomerase from the Japanese arrowroot species *Pueraria lobata*. A similar strategy was employed to synthesize stilbenes by employing the phenylalanine ammonia-lyase from *R. rubra*, a 4-coumarate:CoA ligase from *Lithospermum erythrorhizon*, and the PKS stilbene synthase from *Arachis hypogaea*. When supplied with tyrosine or phenylalanine, *E. coli* cells expressing these three genes produced resveratrol and pinosylvin, respectively. When other carboxylic acids were used as substrates, a number of non-natural stilbenes were produced. To further diversify the products, a promiscuous pinosylvin methyltransferase was introduced to produce a collection of dimethylated stilbenes. This strategy was extended by the addition or substitution of various enzymes in PKS precursor synthesis and/or post-polyketide modification to create a total of 128 polyketide products, 42 of which had not been previously reported [15, 16].

Combinatorial biosynthesis has been performed for the production of terpenoids, the class of chemicals to which the drugs artemisinin and paclitaxel belong. Synthesis of carotenoids has been demonstrated in *E. coli* by combining carotenoid pathway genes from various sources to create 29 different compounds, ten of which had not been isolated from natural sources. To achieve significant yields, the host's metabolism was modified by overexpression of the precursor-generating enzymes 1-deoxy-D-xylulose 5-phosphate synthase, 1-deoxy-D-xylulose 5-phosphate reductoisomerase, and isopentenyl pyrophosphate synthase [38].

While *E. coli* is the most common host used for combinatorial biosynthesis, extensive work on lipopeptides related to the antibiotic daptomycin has been performed in *Streptomyces* hosts [2]. In both *Streptomyces roseosporus* and *Streptomyces fradiae*, the native producers

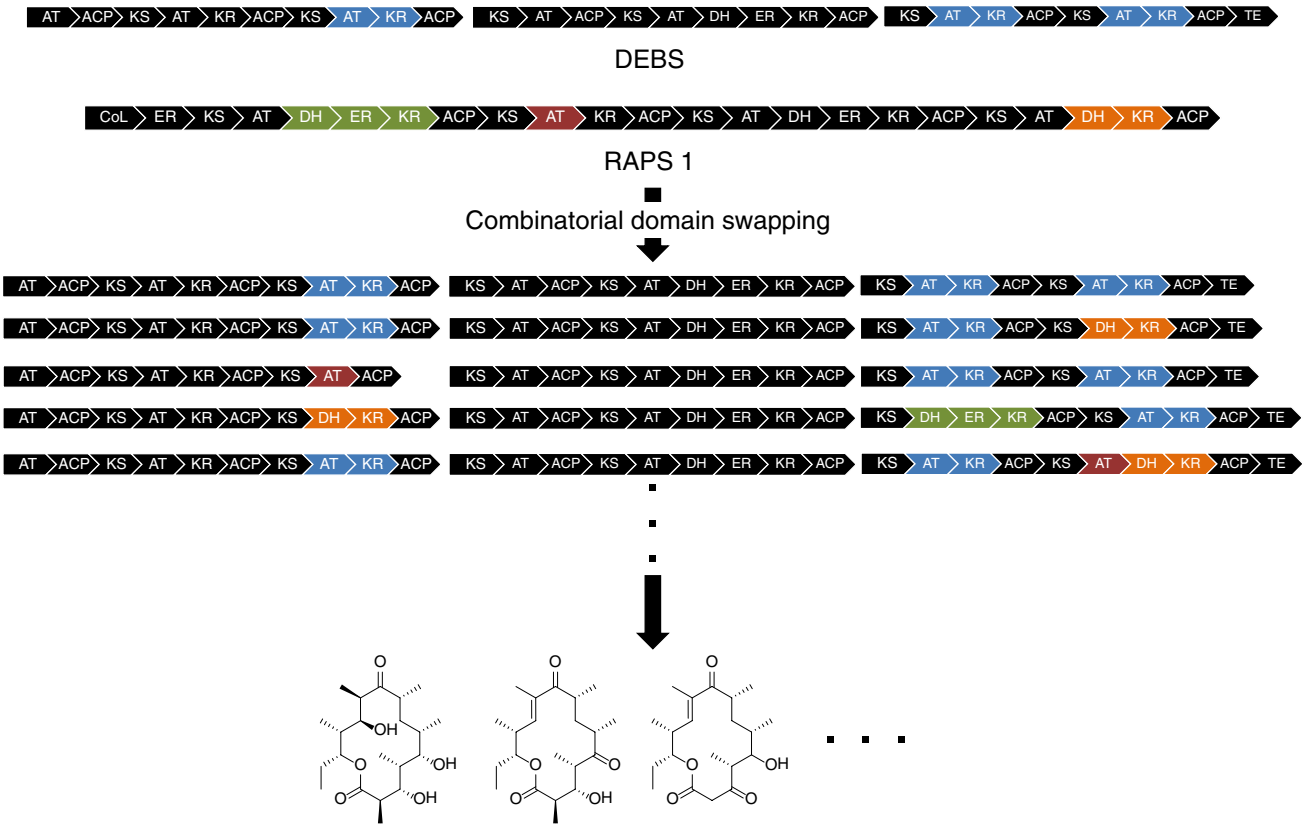

Fig. 1 Combinatorial domain swapping from the rapamycin gene cluster (RAPS from *Streptomyces hygroscopicus*) to the erythromycin gene cluster (DEBS from *Saccharopolyspora erythraea*) yields a library of erythromycin derivatives. Domain abbreviations: *AT* acyltransferase; *ACP* acyl carrier protein; *KS* ketosynthase; *KR* ketoreductase; *DH* dehydratase; *ER* enoylreductase; *TE* thioesterase; *CoL* CoA-ligase

of lipodepsipeptides A21978C and A54145, respectively, genetic modifications including gene deletion, gene replacement, NRPS domain substitution, and module fusion have been introduced. As a result, over 120 novel compounds were produced, some of which exhibited improved therapeutic properties relative to their parent natural products.

Some enzymes with desirable properties from higher organisms, such as fungal membrane-bound cytochrome P450 enzymes, do not express well or are inactive in bacterial hosts. This has prompted adoption of alternate strategies to achieve some reactions, for instance the production of genistein from tyrosine by co-culture of *E. coli* and *Saccharomyces cerevisiae* [15, 16]. However, such co-culture strategies require significant fine-tuning that must be performed on a compound-by-compound basis. Other approaches forgo bacterial hosts altogether in favor of complete production in yeast or fungal cells. A method for combinatorially assembling gene cassettes in yeast artificial chromosomes in vitro has been demonstrated by the creation of a library of flavonoid-producing pathways [32].

Synthetic biology techniques for DNA assembly

The concept of combinatorial biosynthesis is well established through several pioneering examples. Nevertheless, it remains that, in practice, most studies in this field are not rigorously combinatorial. Instead, they are limited to a few selected elements from a much larger library of possibilities. To date, all successful attempts at combinatorial biosynthesis have focused on single pathways or a small set of enzymes, since such a limited set is tractable for laboratories working with standard DNA manipulation techniques. Traditional restriction digestion and ligation-based cloning methods are tedious, time-consuming, and typically require specific tailoring to the entity of interest. Further, many pathways of interest for combinatorial biosynthesis are comprised of several genes and regulatory elements, necessitating a lengthy series of sub-cloning steps en route to the desired pathway constructs. As a result, traditional approaches are not readily amenable to the rapid combinatorial library assembly necessary to create sufficient novel chemical entities for the purpose of drug discovery. The lack of facile, highly efficient manipulation techniques for

libraries of interchangeable genetic elements has heretofore stood as a significant hurdle to true combinatorial biosynthesis. In recent years, however, a number of revolutionary techniques have been developed, transforming arduous constructions into routine tasks.

Modern DNA assembly techniques can broadly be classified into two groups: those based on homology and those based on ligation. Homology-based methods require neighboring DNA fragments to share identical sequences, such that splicing can occur either by annealing and extension of the homologous ends in vitro or by homologous recombination in vivo. Perhaps the most prominent in vitro technique is the one-pot isothermal assembly pioneered by Gibson and coworkers, colloquially known as "Gibson assembly" [12]. In this process, DNA fragments with homologous termini are spliced via three enzymatic reactions. First, T5 exonuclease catalyzes "chew-back" (single-strand degradation) of the 5′ ends of each fragment. This exposes their complementary single-stranded 3′ ends, which anneal to each other in the desired order to form the target construct with single-stranded gaps (Fig. 2). Phusion polymerase then fills in the gaps, and Taq ligase seals the nicks to produce the intact final product, which can subsequently be used to transform a host of choice. A variety of previously developed in vitro assembly techniques present variations on this theme, including sequence and ligase independent cloning (SLIC), which utilizes T4 DNA polymerase for both 3′ chew-back and partial gap-filling, but requires addition of a single deoxynucleotide to switch between the two functions [25]; polymerase incomplete primer extension (PIPE) cloning, which relies on incomplete primer extension during PCR of each fragment to leave single-stranded 3′ ends [20]; and uracil-specific excision reagent (USER) cloning, which utilizes uracil-containing primers and a uracil-specific glycosylase and endonuclease to generate defined single-stranded 3′ ends [4] (Fig. 2). Note that besides Gibson assembly, none of the above methods employ a ligase enzyme, instead requiring nick-sealing to occur in vivo following transformation into the desired host. Further, both SLIC and PIPE cloning also require additional gap-filling in vivo to generate the nicked target construct.

Published shortly after Gibson assembly, circular polymerase extension cloning (CPEC) presents an alternative to the "chew back and anneal" strategy. Starting from a set of DNA fragments with homologous ends, this method instead relies on cycles of heating to denature the duplex fragments, cooling to anneal neighboring strands at their overlapping ends, and polymerase-mediated extension to generate the concatenated duplex. After several cycles, the nicked target construct is formed, which can be sealed in vivo [36]. Site-specific recombination-based tandem assembly (SSRTA), on the other hand, employs the *Streptomyces* phage φBT1 integrase to splice neighboring fragments in vitro [46] (Fig. 2). This method requires each fragment to be flanked by a set of orthogonal recombination sites, and consequently leaves interstitial scar sequences. While this is a clear disadvantage compared to the other techniques described, the absence of a polymerase extension step and the high specificity of the φBT1 integrase for its cognate recognition sequences make this method attractive by avoiding the introduction of mutations or off-target recombination events.

An alternative to in vitro assembly is to allow fragment splicing by native cellular homologous recombination machinery. A key example of this is the DNA Assembler method, which relies on *S. cerevisiae* to assemble DNA fragments with terminal overlap sequences [41] (Fig. 2). In this approach, the assembly host is simultaneously transformed with individual fragments containing homologous ends. The target construct generated in vivo via homologous recombination can then be selected via an incorporated selection marker. Additional in vivo assembly methods include the Red recombination system, in which homologous recombination in *E. coli* is enhanced through expression of the Redαβ proteins from the lambda prophage or RecET proteins from the Rac prophage [48]; mating-assisted genetically integrated cloning (MAGIC), which employs bacterial conjugation to transfer a donor plasmid to the assembly host strain containing a receiver plasmid, a homing endonuclease (to generate linear fragments), and inducible lambda recombinases (to facilitate homologous recombination) [24]; transformation-associated recombination (TAR) cloning, which can be used to clone large portions of genomic DNA via simultaneous transformation of *S. cerevisiae* with genomic DNA containing the target sequence and a receiver vector [27]; and RecET-mediated "direct" cloning, which is similar to TAR cloning but carried out in *E. coli* with inducible expression of the RecET recombinases and requires prior digestion of the genomic DNA to liberate the target sequence as a linear fragment [10]. An ex vivo recombination-based method named for its seamless ligation cloning extract (SLiCE) has also been recently described, which utilizes *E. coli* extracts rather than whole cells to catalyze fragment assembly [49].

Although numerous powerful homology-based assembly methods have been developed, there are still non-trivial limitations to their general utility. Among these is the necessity to avoid multiple fragments with similar homologous ends in the assembly design, as this can lead to incorrect pairing of fragments not intended to be neighbors. Such concerns become significant in clusters containing repeated similar elements, such as the domains of a modular PKS or NRPS or the exogenous promoters and terminators used in pathway refactoring. Thus, there still exists a need for assembly techniques that do not rely on homologous

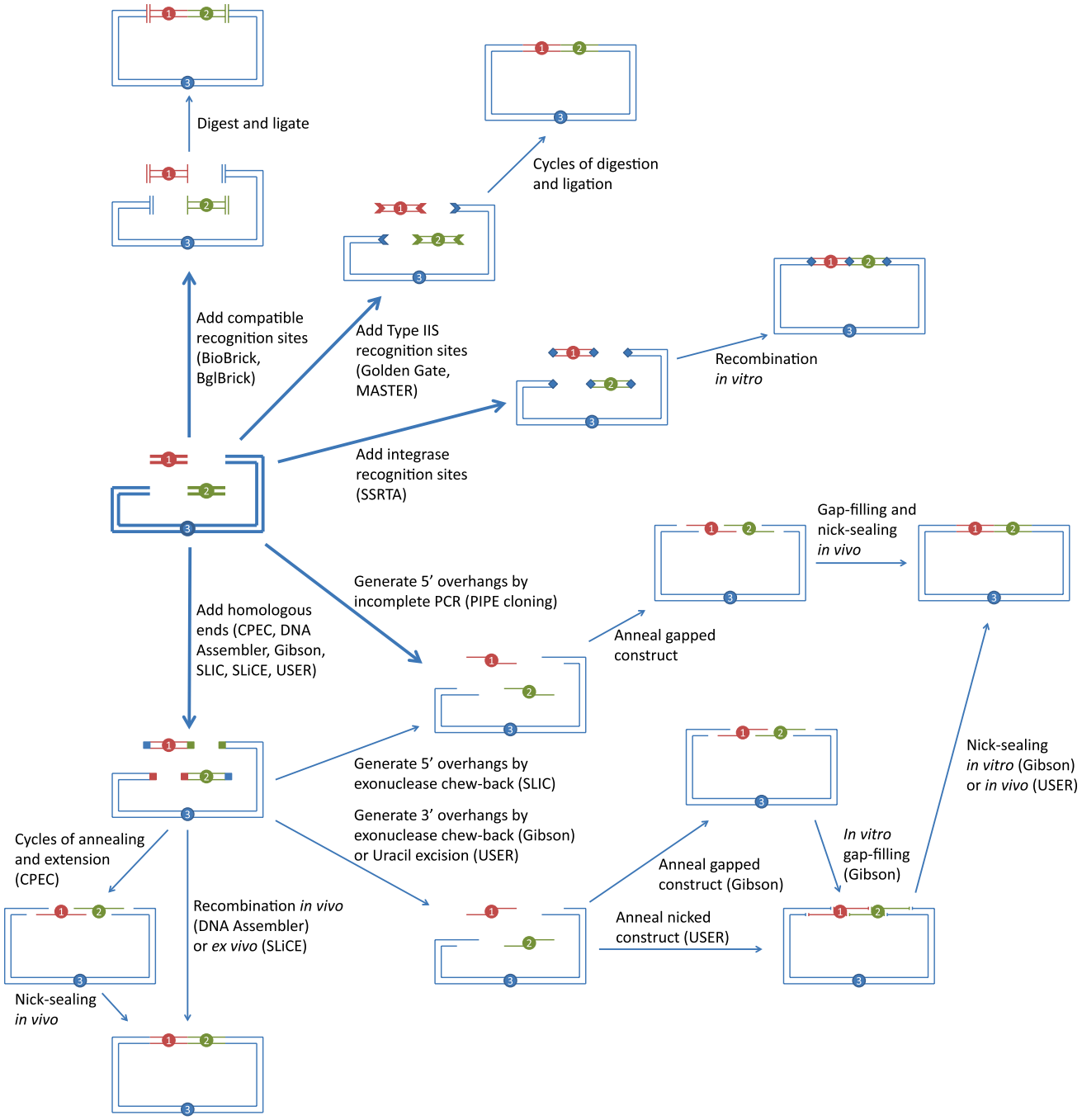

Fig. 2 Modern techniques for multi-component DNA assembly

recombination. Of course, the classic restriction digestion/ligation method is one such technique which generates only short single-stranded overhangs at specific sites. As noted above, however, this method has limited applicability for rapid combinatorial assembly. To facilitate and streamline its application, the concept of BioBricks (and subsequent variants, including BglBricks) has been proposed [42]. BioBrick assembly can be seen as the standardization of traditional cloning techniques. The BioBrick assembly standard dictates the restriction enzyme recognition sequences that should be positioned at the 5′ and 3′ ends of the assembly fragments. Utilization of two restriction enzymes with different recognition sequences but identical single-stranded overhangs (e.g., *Xba*I and *Spe*I) renders the assembly of two fragments an idempotent operation. In other words, correct ligation of two fragments abolishes the recognition sites between them while retaining those at the 5′ and 3′ termini of the product fragment. Thus, the product fragment

can be employed in subsequent assemblies under the same standardized conditions, eliminating the need to identify new restriction enzymes for each fragment in the target construct. Nevertheless, this restriction enzyme-recycling approach necessitates a stepwise rather than simultaneous assembly scheme as only two fragments can be joined per round of assembly. As a result, construction of large secondary metabolite gene clusters by this approach can still be time-consuming.

To reconcile the assembly of several fragments with a convenient enzyme-recycling methodology, a restriction enzyme that can recognize only a single defined sequence but generate many different single-stranded overhangs is needed. Fortuitously, both of these properties are manifested in Type IIS restriction endonucleases, which can bind only to a specific recognition site but cut indiscriminately at a prescribed distance from this site. Thus, by incorporating Type IIS restriction sites at the termini of each fragment, user-defined overhangs can be generated such that simultaneous assembly of multiple fragments in the desired configuration can be achieved. This technique, initially proposed by Engler et al. [8] is termed golden gate assembly.

Note that modern ligation-based cloning techniques still carry with them a major limitation of traditional cloning; namely, the necessity to remove all DNA recognition sites of the selected restriction endonuclease within the fragments to be assembled. To obviate this requirement, Chen and coworkers recently presented a method to limit restriction endonuclease digestion only to the desired terminal sequences [6]. Their method, methylation-assisted tailorable ends rational (MASTER) ligation, expands the utility of golden gate assembly via utilization of *Msp*JI, a type IIS restriction endonuclease containing 5-methylcytosine in its recognition sequence. Through incorporation of 5-methylcytosine in the primers used to amplify each fragment for assembly, digestion only occurs at the desired terminal locations and not within the fragments where only unmodified cytosine's are present.

Current and future applications in combinatorial biosynthesis

The aforementioned techniques have been recently developed, and as such their current applications have

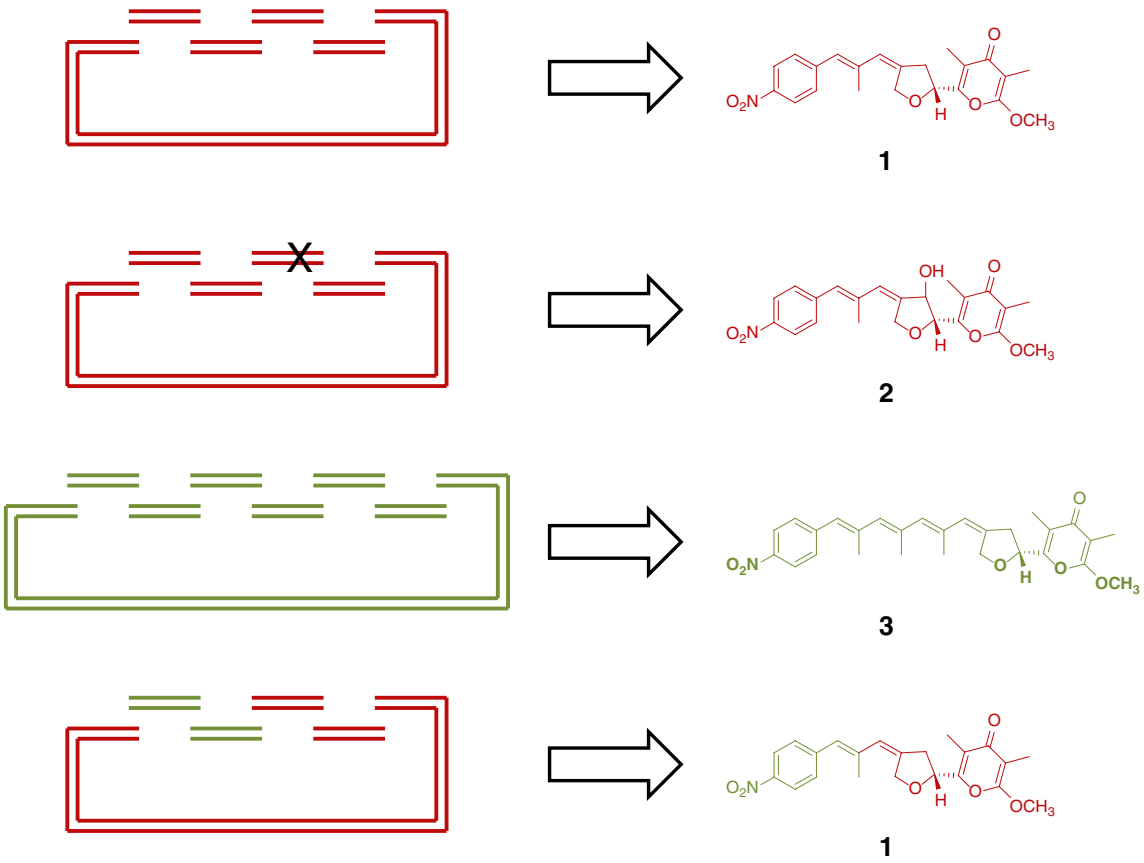

Fig. 3 One step assembly of the aureothin cluster (*red*) by DNA Assembler yields aureothin (**1**) in *S. lividans*. When a single domain is inactivated (depicted with an "*X*"), a derivative (**2**) is produced. Construction of a hybrid gene cluster from the spectinabilin (**3**) pathway (*green*) also yields the expected aureothin product (color figure online)

predominantly been limited to proof-of-concept studies. As with any new technology that seeks to replace established methodologies, a certain degree of recalcitrance to their initial wide-spread adoption is inevitable. Nevertheless, as these techniques continue to gain prominence, more researchers are exploring the scope of their utility, and their potential in the field of combinatorial biosynthesis is just beginning to be realized. An example is the application of the DNA Assembler method for combinatorial biosynthesis of the nitroaryl polyketide aureothin (**1**; Fig. 3) and its derivatives [39, 40]. Modification of the domains of a multimodular PKS (either by mutation or replacement) is a well-established approach in combinatorial biosynthesis for the generation of derivatives of a target polyketide. While it is easy to introduce a mutation to a given sequence via PCR, seamless integration of the mutated PCR product to the full cluster is a non-trivial operation by traditional means. Application of the DNA Assembler method, in contrast, enables facile one-step assembly of the full gene cluster from multiple fragments. As a demonstration, the aureothin cluster was modified by inactivating a dehydratase domain via point mutations in conserved motifs. The mutated fragments were then combined with the remaining fragments that comprise the 29-kb aureothin cluster, assembled in *S. cerevisiae*, and later integrated into the *Streptomyces lividans* genome to produce a new aureothin derivative (**2**; Fig. 3). An additional example was provided in the assembly of a hybrid aureothin pathway comprised of genes from both the native aureothin cluster and the gene cluster for a related nitroaryl polyketide, spectinabilin (**3**; Fig. 3). The hybrid pathway was capable of producing the anticipated aureothin product, again illustrating the versatility of modern assembly techniques for rapidly creating new pathways from disparate sources (Fig. 3).

The DNA assembler method is, of course, not the only technique applicable to rapid assembly of large polyketide synthase-containing gene clusters. One example of an in vitro recombination-based assembly technique applied to construct a natural product pathway was provided by Zhang et al. [46], who utilized SSRTA to assemble the complete 56-kb epothilone gene cluster from *Sorangium cellulosum* So0157-2. They performed the assembly in two steps, first assembling the large PKS *epoD* from four fragments (plus a receiving vector), and then the full cluster from six fragments (plus a receiving vector). While no variations to the gene cluster were introduced in this experiment, an analogous assembly using fragments with site-specific mutations would not be difficult to envision.

A key benefit of modern assembly techniques is the unprecedented scope of the manipulations they facilitate, from individual point mutations to mega base assemblies [11]. Perhaps the most significant contribution of these modern techniques to combinatorial biosynthesis, though, will be to make it truly combinatorial; that is, to enable the facile, simultaneous assembly of many pathways from libraries of interchangeable elements. Early efforts in this direction have already demonstrated this capability. For example, Merryman and Gibson [31] carried out a proof-of-concept experiment in which Gibson assembly was used to join three fragments (two bar-coded open reading frames and a receiver vector). With 79 possibilities for each ORF, a library size of 6,241 was expected, ~92 % of which were identified from Illumina Solexa sequencing. A further example of combinatorial Gibson assembly is the reconstruction of an acetate utilization pathway in *E. coli* [37]. Here, four variants each of the *ackA* and *pta* genes were included in the assembly, along with three possible promoters for each gene, giving a library size of 144. Thirty strains capable of acetate utilization were analyzed, and 10 % of the possible combinations were observed.

Combinatorial assembly of pathway libraries using in vivo techniques has also been recently demonstrated. For example, pathways for xylose utilization and cellobiose utilization in *S. cerevisiae* have been constructed from libraries of promoters via DNA Assembler, dubbed the customized optimization of metabolic pathways by combinatorial transcriptional engineering (COMPACTER) method [7, 45]. An analogous approach was utilized to assemble combinatorial libraries of enzyme variants in the xylose utilization pathway [19]. Eight, ten, and six variants each were included for the three genes in the pathway, respectively, yielding a library size of 480 possible pathways. Twenty-eight isolated pathways from two independent libraries were sequenced and found to all have different gene combinations, demonstrating the unbiased nature of this assembly approach.

At present, the above examples are limited to primary metabolism, and more specifically to the optimization of a single target pathway. However, the extension of these techniques to secondary metabolism can easily be envisioned. In principle, one could assemble libraries of PKS or NRPS modules, for example, to simultaneously generate libraries of derivatives as easily as one can assemble a single pathway. Inclusion of tailoring enzyme libraries with differing stereo- or regio specificities or glycosyl transferases with varied sugar specificities could further diversify the library of compounds generated. Ultimately, a platform could even be developed in which an entire library of derivatives is simultaneously synthesized and screened for a desired bioactivity in a high-throughput format.

Moving forward, the simplification and standardization of modern assembly technologies will make feasible the automation of assembly protocols. Already in silica tools have been developed to design assembly schemes for a variety of methods [14, 26]. Integration of such tools with modern robotic laboratory automation platforms will

enable combinatorial assemblies to be carried out on an unprecedented scale. With a wealth of new pathways generated, new expression hosts will also be needed to maximize production and detection capabilities, such as the versatile genome-minimized *S. avermitilis* SUKA strains [17, 21, 22].

Conclusions

As modern DNA assembly techniques continue to grow in reputation and application, the field of combinatorial biosynthesis is poised for a new generation of growth and innovation. Of course, DNA assembly is not the only obstacle to next-generation combinatorial biosynthesis. For example, difficulties in heterologous expression of enzymes from vastly different species in a desired host can be formidable, although significant strides have been taken to mitigate this difficulty [13]. Further, issues of compatibility between enzymes in designed combinatorial pathways, such as differences in kinetics, localization within the cell, and substrate specificity (or promiscuity) are still significant. Nevertheless, the need for reliable, facile assembly tools remains significant regardless of these concerns, and their development allows these issues to be addressed. The ease with which libraries of new pathways can now be assembled greatly expands the scope of combinatorial strategies, and this potential in the area of natural products is only just beginning to be realized. The number of known and characterized secondary metabolite gene clusters is rapidly increasing through microbial genome mining efforts, further equipping researchers with the requisite diversity of biological parts to harness a truly combinatorial approach and generate unprecedented libraries of new and interesting compounds.

Acknowledgments We thank the National Institutes of Health (GM077596), the National Academies Keck Futures Initiative on Synthetic Biology, and the Energy Biosciences Institute for financial support in our development and application of DNA assembly technologies. REC acknowledges the Drickamer Fellowship support from the Department of Chemical and Biomolecular Engineering at the University of Illinois.

References

1. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215(3):403–410
2. Baltz RH (2012) Combinatorial biosynthesis of cyclic lipopeptide antibiotics: a model for synthetic biology to accelerate the evolution of secondary metabolite biosynthetic pathways. ACS Synth Biol. doi:10.1021/sb3000673
3. Bentley SD, Chater KF, Cerdeno-Tarraga AM, Challis GL, Thomson NR, James KD, Harris DE, Quail MA, Kieser H, Harper D, Bateman A, Brown S, Chandra G, Chen CW, Collins M, Cronin A, Fraser A, Goble A, Hidalgo J, Hornsby T, Howarth S, Huang CH, Kieser T, Larke L, Murphy L, Oliver K, O'Neil S, Rabbinowitsch E, Rajandream MA, Rutherford K, Rutter S, Seeger K, Saunders D, Sharp S, Squares R, Squares S, Taylor K, Warren T, Wietzorrek A, Woodward J, Barrell BG, Parkhill J, Hopwood DA (2002) Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). Nature 417(6885):141–147
4. Bitinaite J, Rubino M, Varma KH, Schildkraut I, Vaisvila R, Vaiskunaite R (2007) USER friendly DNA engineering and cloning method by uracil excision. Nucleic Acids Res 35(6):1992–2002
5. Bleicher KH, Bohm HJ, Muller K, Alanine AI (2003) Hit and lead generation: beyond high-throughput screening. Nat Rev Drug Discov 2(5):369–378
6. Chen WH, Qin ZJ, Wang J, Zhao GP (2013) The MASTER (methylation-assisted tailorable ends rational) ligation method for seamless DNA assembly. Nucleic Acids Res 41(8):e93
7. Du J, Yuan Y, Si T, Lian J, Zhao H (2012) Customized optimization of metabolic pathways by combinatorial transcriptional engineering. Nucleic Acids Res 40(18):e142
8. Engler C, Kandzia R, Marillonnet S (2008) A one pot, one step, precision cloning method with high-throughput capability. PLoS ONE 3(11):e3647
9. Fedorova ND, Moktali V, Medema MH (2012) Bioinformatics approaches and software for detection of secondary metabolic gene clusters. Methods Mol Biol 944:23–45
10. Fu J, Bian X, Hu S, Wang H, Huang F, Seibert PM, Plaza A, Xia L, Muller R, Stewart AF, Zhang Y (2012) Full-length RecE enhances linear–linear homologous recombination and facilitates direct cloning for bioprospecting. Nat Biotechnol 30(5):440–446
11. Gibson DG, Benders GA, Andrews-Pfannkoch C, Denisova EA, Baden-Tillson H, Zaveri J, Stockwell TB, Brownley A, Thomas DW, Algire MA, Merryman C, Young L, Noskov VN, Glass JI, Venter JC, Hutchison CA 3rd, Smith HO (2008) Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. Science 319(5867):1215–1220
12. Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA 3rd, Smith HO (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods 6(5):343–345
13. Gustafsson C, Minshull J, Govindarajan S, Ness J, Villalobos A, Welch M (2012) Engineering genes for predictable protein expression. Protein Expr Purif 83(1):37–46
14. Hillson NJ, Rosengarten RD, Keasling JD (2012) j5 DNA assembly design automation software. ACS Synth Biol 1(1):14–21
15. Horinouchi S (2008) Combinatorial biosynthesis of non-bacterial and unnatural flavonoids, stilbenoids and curcuminoids by microorganisms. J Antibiot (Tokyo) 61(12):709–728
16. Horinouchi S (2009) Combinatorial biosynthesis of plant medicinal polyketides by microorganisms. Curr Opin Chem Biol 13(2):197–204
17. Ikeda H, Shin-Ya K, Omura S (2013) Genome mining of the *Streptomyces avermitilis* genome and development of genome-minimized hosts for heterologous expression of biosynthetic gene clusters. J Ind Microbiol Biotechnol
18. Khaldi N, Seifuddin FT, Turner G, Haft D, Nierman WC, Wolfe KH, Fedorova ND (2010) SMURF: genomic mapping of fungal secondary metabolite clusters. Fungal Genet Biol 47(9):736–741
19. Kim B, Du J, Eriksen DT, Zhao H (2013) Combinatorial design of a highly efficient xylose-utilizing pathway in *Saccharomyces cerevisiae* for the production of cellulosic biofuels. Appl Environ Microbiol 79(3):931–941
20. Klock HE, Koesema EJ, Knuth MW, Lesley SA (2008) Combining the polymerase incomplete primer extension method for cloning and mutagenesis with micro screening to accelerate structural genomics efforts. Proteins 71(2):982–994
21. Komatsu M, Komatsu K, Koiwai H, Yamada Y, Kozone I, Izumikawa M, Hashimoto J, Takagi M, Omura S, Shin-Ya K, Cane DE,

Ikeda H (2013) Engineered *Streptomyces avermitilis* host for heterologous expression of biosynthetic gene cluster for secondary metabolites. ACS Synth Biol 2(7):384–396
22. Komatsu M, Uchiyama T, Omura S, Cane DE, Ikeda H (2010) Genome-minimized *Streptomyces* host for the heterologous expression of secondary metabolism. Proc Natl Acad Sci USA 107(6):2646–2651
23. Li JW-H, Vederas JC (2009) Drug discovery and natural products: end of an era or an endless frontier? Science 325(5937):161–165
24. Li MZ, Elledge SJ (2005) MAGIC, an in vivo genetic method for the rapid construction of recombinant DNA molecules. Nat Genet 37(3):311–319
25. Li MZ, Elledge SJ (2012) SLIC: a method for sequence- and ligation-independent cloning. Methods Mol Biol 852:51–59
26. Linshiz G, Stawski N, Poust S, Bi C, Keasling JD, Hillson NJ (2013) PaR–PaR laboratory automation platform. ACS Synth Biol 2(5):216–222
27. Ma H, Kunes S, Schatz PJ, Botstein D (1987) Plasmid construction by homologous recombination in yeast. Gene 58(2–3):201–216
28. Marchler-Bauer A, Zheng C, Chitsaz F, Derbyshire MK, Geer LY, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Lu S, Marchler GH, Song JS, Thanki N, Yamashita RA, Zhang D, Bryant SH (2013) CDD: conserved domains and protein three-dimensional structure. Nucleic Acids Res 41:D348–D352 (Database issue)
29. McDaniel R, Thamchaipenet A, Gustafsson C, Fu H, Betlach M, Ashley G (1999) Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel "unnatural" natural products. Proc Natl Acad Sci USA 96(5):1846–1851
30. Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA, Weber T, Takano E, Breitling R (2011) AntiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Res 39:W339–W346 (Web Server issue)
31. Merryman C, Gibson DG (2012) Methods and applications for assembling large DNA constructs. Metab Eng 14(3):196–204
32. Naesby M, Nielsen SV, Nielsen CA, Green T, Tange TO, Simon E, Knechtle P, Hansson A, Schwab, Titiz O, Folly C, Archila RE, Maver M, van Sint Fiet S, Boussemghoune T, Janes M, Kumar AS, Sonkar SP, Mitra PP, Benjamin VA, Korrapati N, Suman I, Hansen EH, Thybo T, Goldsmith N, Sorensen AS (2009) Yeast artificial chromosomes employed for random assembly of biosynthetic pathways and production of diverse compounds in *Saccharomyces cerevisiae*. Microb Cell Fact 8:45
33. Newman DJ, Cragg GM (2012) Natural products as sources of new drugs over the 30 years from 1981 to 2010. J Nat Prod 75(3):311–335
34. Ohnishi Y, Ishikawa J, Hara H, Suzuki H, Ikenoya M, Ikeda H, Yamashita A, Hattori M, Horinouchi S (2008) Genome sequence of the streptomycin-producing microorganism *Streptomyces griseus* IFO 13350. J Bacteriol 190(11):4050–4060
35. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer EL, Eddy SR, Bateman A, Finn RD (2012) The Pfam protein families database. Nucleic Acids Res 40:D290–D301 (Database issue)
36. Quan J, Tian J (2011) Circular polymerase extension cloning for high-throughput cloning of complex and combinatorial DNA libraries. Nat Protoc 6(2):242–251
37. Ramon A, Smith H (2011) Single-step linker-based combinatorial assembly of promoter and gene cassettes for pathway engineering. Biotechnol Lett 33(3):549–555
38. Sandmann G (2002) Combinatorial biosynthesis of carotenoids in a heterologous host: a powerful approach for the biosynthesis of novel structures. ChemBioChem 3(7):629–635
39. Shao Z, Luo Y, Zhao H (2011) Rapid characterization and engineering of natural product biosynthetic pathways via DNA assembler. Mol BioSyst 7(4):1056–1059
40. Shao Z, Zhao H (2012) DNA assembler: a synthetic biology tool for characterizing and engineering natural product gene clusters. Methods Enzymol 517:203–224
41. Shao Z, Zhao H, Zhao H (2009) DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. Nucleic Acids Res 37(2):e16
42. Shetty RP, Endy D, Knight TF Jr (2008) Engineering BioBrick vectors from BioBrick parts. J Biol Eng 2:5
43. Wong FT, Khosla C (2012) Combinatorial biosynthesis of polyketides—a perspective. Curr Opin Chem Biol 16(1–2):117–123
44. Yoon YJ, Beck BJ, Kim BS, Kang HY, Reynolds KA, Sherman DH (2002) Generation of multiple bioactive macrolides by hybrid modular polyketide synthases in *Streptomyces venezuelae*. Chem Biol 9(2):203–214
45. Yuan Y, Du J, Zhao H (2013) Customized optimization of metabolic pathways by combinatorial transcriptional engineering. Methods Mol Biol 985:177–209
46. Zhang L, Zhao G, Ding X (2011) Tandem assembly of the epothilone biosynthetic gene cluster by in vitro site-specific recombination. Sci Rep 1
47. Zhang W, Tang Y (2008) Combinatorial biosynthesis of natural products. J Med Chem 51(9):2629–2633
48. Zhang Y, Buchholz F, Muyrers JP, Stewart AF (1998) A new logic for DNA engineering using recombination in *Escherichia coli*. Nat Genet 20(2):123–128
49. Zhang Y, Werling U, Edelmann W (2012) SLiCE: a novel bacterial cell extract-based DNA cloning method. Nucleic Acids Res 40(8):e55