DNA fabrication of genetic cassettes at base-level precision is transforming genetic engineering from a laborious art to an information-driven discipline. Although substantial advances have been made in the development of DNA fabrication, the methods employed vary widely based on the length of the DNA. All of these methods are available commercially, but can also be performed at the molecular biology bench using typical reagents and procedures. Because the technology is not mature and is still evolving rapidly, it is helpful to gain some understanding of the different steps in this process and the associated technical challenges to successfully take advantage of DNA fabrication in a research project.

Introduction
The de novo fabrication of custom DNA molecules is a transformative technology that will radically affect the biotechnology industry. Even before the foundations of molecular biology were established, pioneers were working on the chemical synthesis of genes. Box 1 provides a short history of gene synthesis (see Glossary). Basic genetic engineering techniques for manipulating DNA in vitro opened an incredible breadth of opportunity in the life sciences and biotechnology industries. However, genetic engineering has now moved beyond the introduction of single genes into cells. Multi-gene cassettes are now common and progression towards whole genome engineering is rapid. The synthesis of DNA molecules has resurfaced as a time- and cost-limiting step in genetic engineering.

DNA synthesis demystified

Theoretically, DNA fabrication methods that are rooted in chemical synthesis could transform synthesis into a generic, predictable and scalable process enabling the generation of any DNA. By liberating the process from the confines of pre-existing sequences, the problem of composition and design becomes orthogonal to the problem of physical construction. In other words, the question becomes how does one design something useful once complete creative freedom is realized? Therefore, as gene synthesis becomes a commodity, biologists will spend more time designing custom DNA molecules and characterizing their performance, and less time constructing them. Examples of this creative freedom are already being achieved through the creation of new proteins with novel enzymatic activities [3,4].

Glossary

BioBricks: functional units of genetic composition flanked by a standardized polylinker enabling their assembly using a singular standard assembly reaction.
DNA fabrication: any method of synthesizing large DNA at base-level precision.
Endonuclease: an enzyme that cleaves DNA internally.
Exonuclease: an enzyme that degrades DNA from its termini.
Gene synthesis: synthesis of gene-length DNA from chemically derived oligonucleotides.
LBS: Ligation by selection, a method of assembling synths using vectors with different antibiotic resistance markers.
LCR: Ligase chain reaction, a method of assembling many oligonucleotides into a longer double-stranded DNA employing a thermostable DNA ligase.
LIC: Ligation-independent cloning, a method for cloning PCR products into vectors using annealing of long single-stranded termini.
Oligonucleotide synthesis: chemical synthesis of single-stranded DNA typically between 10 and 100 bp in length.
PCA: Polymerase chain assembly, a method of assembling many oligonucleotides into a longer double-stranded DNA employing a thermostable polymerase.
PCR: Polymerase chain reaction, a method for specifically amplifying DNA using a thermostable DNA polymerase and synthetic oligonucleotides.
Phosphoramidite chemistry: Solid-phase reaction of the 5’-OH of a growing DNA chain with the 3’ phosphoramidite moiety of a free deoxynucleotide.
SLIC: Sequence- and ligation-independent cloning, a method for joining multiple DNA fragments in a single reaction using in vitro homologous recombination and annealing of single-stranded termini generated through the exonuclease activity of T4 DNA polymerase.
SOEing: Gene splicing by overlap extension, a method for joining two PCR products into one employing short homologous sequences.
Synthon: Fragment of a large synthetic DNA, typically 500 bp to 1 kb in length, constructed by gene synthesis.
TBIO: Thermodynamically balanced inside out synthesis. This is an extension protocol that is initiated at the center of the synthetic gene.
Type I restriction endonuclease: Enzyme that binds sequence-specifically to a site on DNA and cuts adjacent to it.
UDG: Uracil DNA glycosidase, an enzyme that catalyses the excision of uracil base incorporated into DNA.
UrACl excision-based cloning: a ligase-free cloning method based on UDG-containing uracil-specific excision reagent.
Box 1. A brief history of DNA synthesis

The idea of building synthetic genes using a chemical process can be traced back to 1965. At the time, the gene encoding an alanine tRNA was the only tRNA sequence known. Khorana and his group embarked on the daring project to synthesize its 77 nucleotides [40] and it took them five years to synthesize the required 17 oligonucleotides using chemical synthesis processes that were still in their infancy [41,42]. These oligonucleotides were then later assembled using the recently discovered ligase [43]. Seven years later, the gene encoding the 14 amino acid hormone somatostatin was synthesized and expressed in Escherichia coli [44]. Its gene was synthesized by ligation of eight oligonucleotides varying in length from 11 to 16 bases. Improvements over the next ten years led to the synthesis of DNA fragments up to 1 kb [5].

The invention of the PCR in the mid 1980s [45,46] enabled the development of new fabrication methods that relied on an amplification step. PCR also reduced the cost of DNA synthesis, increased productivity and increased the length of available synthetic DNA molecules. Nothing illustrates this trend better than the 2004 synthesis of the 32-kb gene cluster, which encodes one of the largest known protein complexes, erythromycin polypeptide synthase (PKS) [14]. Although the length of this synthesis has been surpassed, it still demonstrates the ability to create functional proteins through gene synthesis. Gene synthesis to optimize heterologous expression of proteins is frequently being carried out and >500 genes from 200 papers are referenced in the Synthetic Gene Database (http://www.evolvingcode.net/codon/sgdbin/dex.php) [47]. As the fields of protein engineering and gene synthesis more tightly intertwine, we expect to see an increase in the number of genes for completely novel proteins with entirely new protein functions synthesized [4].

Since 2000, the field has moved from gene synthesis to whole genome synthesis. In 2002, the 7.5-kb poliovirus genome was synthesized from oligonucleotides [6]. Using an in vitro gene expression system, this synthetic DNA molecule produced viral particles that proved partially virulent. More recently, the genome of the poliovirus has been redesigned by systematically using rare pairs of codons. The resulting virus is poorly virulent and could be used as a vaccine [6]. In 2003, the 5.4-kb genome of the bacteriophage4x174 was assembled in 14 days [8]. After electroporation in the phage host, fully infective virions were recovered. The strain of the influenza virus responsible for the pandemic that killed an estimated 20–50 million people worldwide was resurrected synthetically in 2005 and characterized in animal experiments [48]. The virus genome was sequenced from samples collected on remains of a victim buried in the Alaskan permafrost. Genome synthesis is now moving beyond short viral genomes. The synthesis of the 583 kb of the Mycoplasma genitalelum genome has been recently reported [30]. Sequence analysis proved that it is structurally correct but it has not been established that this genome is functional.

This idealistic vision of large scale, affordable gene synthesis has not yet materialized in spite of rapid progress. This review provides a brief description of current DNA fabrication methods as they are used today, paired with a discussion of new and potential future methods that aim to improve accessibility and widen use. Notwithstanding, care should be taken by the community of scientists participating in gene synthesis. Wider accessibility of DNA synthesis now presents the research community with the challenge of mitigating biosecurity risks, as discussed in Box 2.

DNA fabrication

At its core, DNA fabrication relies on base level synthesis of DNA oligomers. The essential feature of DNA fabrication is that no naturally isolated DNA is used. Although clonal plasmid-based intermediates might exist during the assembly of a target DNA, every base originated as a phosphoramidite molecule at the beginning of the process. Today, all fabrication methods begin with solid-phase phosphoramidite chemistry to construct single-stranded DNA between 10 and 100 base pairs (bp) long, which are enzymatically assembled into larger molecules. This process is commonly referred to as ‘gene synthesis’ and can be used to synthesize sequences up to 1 kilobase (kb) long. Still larger target DNA sequences require investigators to assemble partial products into the desired full-length construct. Multi-kilobase DNA can be enzymatically assembled from 1-kb DNA segments, whereas megabase-length DNA requires in vitro recombination methods. The details of DNA fabrication are therefore not monolithic and are distinguishable based on the size of the target DNA.

Assembling oligomers into genes

Gene synthesis is the step during which oligonucleotides (oligos) are combined into DNA fragments of several hundred bases in length. Numerous protocols have been described and extensively reviewed [5]. In this section, we discuss three representative protocols, polymerase chain assembly (PCA), ligase chain reaction (LCR) [6] and thermodynamically balanced inside out synthesis (TBIQ), with the goal of underscoring the important parameters that need to be considered when selecting or developing an assembly protocol.
**Polymerase chain assembly**

The PCA procedure uses oligos spanning both strands of the entire gene sequence in equimolar quantities. Oligos with partial overlaps anneal and are extended such that each becomes longer and can be further extended by hybridizing to other oligos or products of subsequent extensions (Figure 1). Stemmer et al. [7] first used PCA methods to assemble a 900-bp β-lactamase gene. This synthesis used 40-nt oligonucleotides, a reaction mixture containing Taq and Pfu polymerases, and a 55 cycle reaction [8]. Because all the oligos are present in approximately the same quantities, PCA is not an amplification reaction. However, a subsequent PCR based on the two terminal oligos is the final step that is used to amplify only the full-length product.

**Ligase chain reaction**

LCR exploits the special properties of Taq DNA ligase to repair nicks within double-stranded DNA at high temperatures. Similarly to PCR, an LCR involves heating cycles for denaturation, annealing and ligation. The joining of two oligos only occurs when they are properly annealed to a third template oligo. This strict requirement for proper base-pairing between the oligos and stringent annealing conditions selects against oligos with mutations being used in the reaction and thereby reduces the error rate in the final product. Often, LCR is used in conjunction with a final assembly reaction by PCA. The synthesis of the bacteriophage φX174 was done in this manner [8]. Another measure was taken to minimize mutations in this protocol; a polyacrylamide gel electrophoresis (PAGE) purification of the 259 oligonucleotides was used to enrich for full length oligos. Only 50% of the oligos were full length. Because all the oligonucleotides were the same length (42-mers), they could be purified in two strand-specific pools. After PAGE purification, the oligos were not present in equimolar concentrations and PCA was necessary to generate the full-length product. The slow cooling protocol (0.1 °C/s) and a high-fidelity PCR kit helped minimize errors. After PCA and PCR amplification, the linear bacteriophage φX174 DNA was cleaved at both extremities with PstI, circularized by ligation and replicated in bacteria to generate phage particles. Extracts from this cell culture were spread over a bacterial lawn to determine whether viable phages were present by plaque formation. The phage genomes from four plaques were sequenced and one matched the Sanger sequence exactly [8], whereas one, three and five mutations were found in the others.

*Figure 1. PCA assembly of a DNA construct. A target sequence is shown in the top panel of this figure. The different color segments represent the oligos that are synthesized to build the construct. The pool of oligos is assembled in equimolar amounts and allowed to anneal. The annealed oligos are extended in the 3' direction until the end of their partner oligo is reached. The double-stranded DNA is melted and reannealed with extension products and any remaining oligos. Each extension reaction results in progressively longer products and full length products are eventually synthesized. At this step, the terminal oligos are added to the reaction and full length products from the previous reactions are amplified by PCR and subsequently cloned and sequenced.*
**Thermodynamically balanced inside out synthesis**

TBIO synthesis is an extension protocol that is initiated at the center of the synthetic gene and is progressively extended in both directions by using overlapping oligos that are homologous to the forward strand at the 5’ end of the gene and against the reverse strand at the 3’ end of the gene (Figure 2). Three human genes, PKB2 (1494 bp), S6K1 (1622 bp) and PDK1 (1712 bp), were synthesized using TBIO [9]. The procedure starts at the center of the molecule, at which only one pair of oligos from opposite strands overlaps. Each strand is extended and now overlaps with another complementary oligo that will serve as a template for additional extension in the next cycle of the reaction. Rather than using primers to cover both sense and antisense strands, as in PCA, this method uses overlapping forward oligos to elongate to the left and overlapping reverse oligos to elongate to the right. Lengthening occurs progressively and each iteration of the reaction contains four to six pairs of oligos combined along a concentration gradient [9] (Figure 2). After each reaction, the full length product is gel purified and is used as a template for extension in both directions.

**Mitigating synthesis errors**

The authors of the phage synthesis study mentioned earlier estimated that sequences assembled using their protocols would contain approximately two mutations per kilobase if no selection step based on infectivity was included [8]. In this context, it is worth noting that the probability of obtaining a molecule with the specified sequence decreases exponentially as the length of the sequence increases. Various methods have therefore been developed to improve the fidelity of oligonucleotide assembly. The ligase chain reaction used for phage synthesis, described earlier, incorporated gel purification of oligos, stringent annealing conditions and ligation as methods to decrease the probability of introducing errors. Other methods are also described here.

Bang and Church [10] devised a one-step synthesis method that minimized errors in the assembly of the 1-kb *Dpo4* gene from 48 oligos. The DNA design used ‘guide’ oligos that enabled circularization of the DNA. DNA that was improperly assembled and non-circular was eliminated by exonuclease treatment and circular DNA with mismatches was linearized by a mismatch-cleaving endonuclease and could then be degraded with exonucleases. Using these two enzymatic treatments in conjunction with a stringent annealing temperature, this method resulted in a 0.025% error rate that compares favorably to the 0.183% error rate of a conventional PCA reaction. Nevertheless, the high rate of error incurred during chemical synthesis typically limits the size of a DNA molecule that can be assembled directly from oligonucleotides to approximately 1 kb.

Selective elimination of synthetic DNA sequences with mismatches from further synthesis has also been proposed by using the MutS protein from *Thermus aquaticus* to retard electrophoretic mobility [11]. DNA fragments containing mismatches bind MutS and migrate at a higher molecular weight in polyacrylamide gel electrophoresis than perfectly complementary sequences. The faster migrating DNA fragments that do not have mismatches can be eluted and further assembled. This method seems to decrease error rate to 1/15 of that where *mutS* is not used [11]. Similarly, immobilized MutS has also been used to remove mismatch-containing DNA sequences [12]. DNA synthesis products are denatured and then re-hybridized and the immobilized MutS enriches the unbound population of DNA for error free molecules; this process is referred to as coincidence filtering. For DNA sequences in which the length of the molecule is such that having a completely mismatch-free molecule is highly unlikely, the rehybridized DNA molecules can be digested into fragments with an endonuclease and mismatch containing fragments can be removed. Subsequent PCA reactions regenerate the DNA molecule from the mismatch-reduced pool of fragments in a process referred to as consensus shuffling [12].

Shapiro and coworkers have developed a recursive construction strategy that involves the assembly of oligonucleotides in a pairwise manner rather than the one-pot methods conventionally employed with PCA or LCR [13]. Although the method does not intrinsically change the rate of mutation, it allows for the facile reconstruction of a correct DNA from a set of first-generation products with mutations.

**Cloning synthons**

To assemble error-free fragments into larger DNA molecules, cloning strategies are required that typically involve two steps. In the first step, DNA molecules of several hundred base pairs are produced by various methods, including those mentioned earlier. These segments of DNA, referred to as synthons by Kodumal *et al* [14], are cloned into a plasmid and their sequences are confirmed. These synthons are then combined in a second step into larger sequences. By locking in a single DNA sequence through cloning, mutations introduced during the preliminary stages of DNA fabrication are eliminated before subsequent assembly steps.
For synthesis projects that require more than a few synthons to form the complete sequence, it is inefficient to design a custom cloning strategy for each synthon. One strategy for standardizing the introduction of synthetic DNA into a cloning vector is to use the same set of restriction enzymes and a conventional digestion and ligation strategy. However, such a strategy would require uniqueness of the chosen restriction sites and is, therefore, not a one-method-fits-all solution. A variety of alternative strategies have been developed that do not have this restriction. The TOPO TA cloning kit (http://products.invitrogen.com) from Invitrogen uses a pre-prepared vector in combination with Vaccinia topoisomerase to capture PCR products. The use of a non-proofreading polymerase such as Taq polymerase is necessary for this approach because it employs ligation of the single A overhangs. Although simple, popular and highly efficient, TOPO TA cloning provides no control over the orientation of the PCR product in the final vector and might, therefore, require colony screening for some applications.

An alternative strategy, termed ligation-independent cloning (LIC) [15–17], uses long complementary overhangs to join DNA segments. Because no ligation is necessary, this technique is fast and versatile. Uracil excision-based cloning is one method by which cohesive overhangs can be generated (Figure 3). This method requires PCR amplification with primers that contain a uracil ~10 bp from the end of the amplicon. The PCR product is treated with uracil DNA glycosidase (UDG) and T4 endonuclease V to generate 3’ overhangs that can be annealed with a complementary sequence [18,19]. Commercial kits are available for this technique including New England Biolabs USER Friendly cloning kit (http://www.neb.com/nebecomm/products/productsE5500.asp). The exonuclease activity of T4 DNA polymerase in the presence of a specific dNTP can also be used to generate complementary ends [15–17]. However, these ends must lack the nucleotide present in the reaction mixture [15].

A wide variety of additional strategies exist to rapidly transfer PCR products into cloning vectors using recombination. These include the use of homologous recombination with lambda Red recombinase [20] and the use of site-specific recombination using Gateway cloning with lambda Int recombinase [21] or Cre recombinase [22].

Synthon-joining strategies
With a complete collection of synthons in hand, larger DNA can be assembled. One assembly strategy is to design unique restriction sites at the junctions between the synthons. Restriction sites can be introduced as silent mutations in the coding regions of the sequence [23,24]. This fairly conventional cloning strategy, however, becomes less successful as the sequence length increases. For example, it can be difficult to find unique restriction sites in the synthesized sequence or to introduce silent mutations in non-coding regions. Even in coding regions, codon modification can negatively affect protein translation [25–27]. Furthermore, this approach will require a repertoire of in vitro reactions that would not be automatable, thereby restricting the scalability of the approach.

Ligation by selection assembly
Because of the limitations mentioned, several groups have developed cloning strategies that rely on a standard assem-

---

**Figure 3.** Uracil excision-based cloning to seamlessly join two DNA sequences. A target sequence is synthesized from two partial sequences (synthons) by virtue of a shared junction sequence of approximately eight base pairs (highlighted by the vertical gray bar). The two partial sequences that each contained the junction sequence were synthesized from oligos. Each is PCR amplified using primers in the junction sequence, which contain deoxyuridines (dUs) in place of deoxythymidines (dTs). The dU-containing primers are represented by the red arrows. The resulting amplicons contain dUs on one DNA strand (represented by the red rectangles) and the two synthons contain the dUs on opposite strands. Uracil excision removes the uracil-containing portion of one strand of DNA by treatment of the dU-containing amplicons with a combination of uracil DNA glycosidase (UDG) and T4 endonuclease V (or UDG and DNA glycosylase-lyase endo VIII), which leaves the cohesive ends of complementary sequences on the two amplicons. Because the complementary sequence is approximately 8–10 bp, the annealed DNA is sufficiently stable to be transformed without a ligation step.
bly reaction. The ligation by selection (LBS) method uses only three enzymes (XhoI, BsaI and BbsI) [14,18]. BsaI and BbsI are type II restriction endonucleases that cleave DNA at a defined distance from their non-palindromic, asymmetric recognition sites. This property makes it possible to design cloning vectors having unique sites for these two enzymes without having to introduce the sites in the synthon sequence. Even though minimal enzymes are required, this strategy retains the limitation that the sequences to be assembled must not contain these restriction sites. To facilitate assembly, cloning vectors that rely on four different antibiotics have been designed to avoid the purification of DNA fragments after the restriction digest. The restriction sites are placed on the cloning vectors in such a manner that one synthon can be excised from its cloning vector along with a specific antibiotic resistance gene. This excised fragment is inserted into another synthon-containing vector in the location from which one of the antibiotic resistance genes has been excised with compatible restriction enzymes. This makes it possible to select clones that have the synthons from the two initial plasmids in the desired position using dual antibiotic selection. Using four cloning vectors and three assembly steps, it was possible to generate a 3.4-kb fragment that included eight synthons [14].

**BioBrick™ assembly**

The methods of synthesis described thus far are directed towards a single specific target sequence. The synthons are chosen based on the most efficient strategy for constructing the individual target without regard for the functions they encode. BioBrick-based methods of assembly take a different approach. Instead of synthesizing and assembling synthons, the primary unit in this approach is a ‘basic’ part encoding features such as promoters and coding sequences [28]. BioBricks include basic parts or combinations of parts surrounded by prefix and suffix sequences containing restriction sites, which enable a standard methodology for pairwise joining of BioBricks. The BioBrick approach carries the additional advantage that basic parts can often be re-used in unrelated projects and a vibrant community-shared collection of these biological parts has emerged. The original strategy for BioBrick assembly involved synthesizing basic parts flanked by unique XbaI and SpeI restriction sites (Figure 3). Upon cleavage, these sites generate compatible cohesive ends that can be joined by ligation. The junction between basic parts retains remnants of the joined restriction sites, which are referred to as a scar sequence, and is uncleavable by XbaI and SpeI. Much like the LBS approach, iterative rounds of digestion and ligation result in polymers of the basic parts. The original BioBrick standard produced a scar containing a stop codon in frame with the upstream part. More recently, new standards have emerged including the Biofusion, CRG and BB standards that are designed to produce a scar sequence that does not contain an in-frame stop between the assembled parts and enable the construction of protein fusion parts. Furthermore, a scarless strategy called BioBricks++ employing type II restriction enzymes and blunting reactions has been described. The BioBricks Foundation has taken the role of documenting these technical standards [29]. More information is available at [http://bbf.openwetware.org/Standards/Technical/Formats.html](http://bbf.openwetware.org/Standards/Technical/Formats.html). In both the synthon and BioBrick approaches, traditional molecular biology techniques must be used to assemble this larger DNA, although much effort has been made to simplify, standardize and automate these methods.

**Chew back, anneal and repair**

Gibson et al. [30] developed a procedure similar to the SLIC method [2] to synthesize the *M. genitalium* genome. This method of *in vitro* assembly permits researchers to assemble gene segments without the need to maintain unique restriction sites or to assemble sequences in cells in which the gene product could be toxic or the sequence too long to maintain with fidelity. Segments of the final sequence are synthesized so that there are overlapping sequences between adjacent segments. The gene segments used as the starting material for the *M. genitalium* genome were excised from recombinant plasmids and digested with 3’ exonuclease. The overlapping sequences between adjacent segments became complementary cohesive overhangs that were annealed, extended to fill any gaps and then ligated. This method also does not leave a scar sequence like the BioBricks assembly method because the cohesive ends are generated by nonspecifically chewing back one strand of DNA.

**Homologous recombination**

Construction of large DNA molecules ultimately reaches a scale in which *in vitro* reactions and transformation procedures are no longer viable strategies. Manipulation of DNA on the 20-kb+ scale is therefore best performed by *in vivo* recombination methodologies. These methods employ homologous recombination in *Bacillus subtilis, E. coli* or yeast. Tsuge and coworkers described a ‘domino method’ to assemble a 134.5-kb rice chloroplast DNA in the genome of *B. subtilis* [31]. First, 4–6-kb DNA fragments, termed dominoes, were constructed in conventional pBR322-derived plasmids in *E. coli* and associated with either a chloramphenicol resistance gene or an erythromycin resistance gene. The ends of the DNA contained regions of homology on one end to the *B. subtilis* genome and on the other end to another domino fragment. Using multiple rounds of transformation and selection alternatively with erythromycin or chloramphenicol, the entire rice chloroplast DNA sequence was assembled in the *B. subtilis* genome. Subsequently, the DNA was transferred back into an *E. coli* cloning vector using a homologous recombination strategy. In *E. coli*, in which the natural efficiency of homologous recombination is low, the rate can be increased by expression of the lambda Red recombinase genes. A variety of ‘recombineering’ methods for transferring large DNA between plasmids or between plasmids and the genome have been devised using this system [32–34]. Finally, the intrinsic high efficiency of homologous recombination and ability to replicate large DNA make yeast an attractive organism for similar manipulations, particularly for the synthesis of bacterial genome-length DNA [30].
Synthesis of whole genomes

The first complete bacterial genome synthesized, Mycoplasma genitalium, is the largest DNA synthesis accomplished to date, with a size of 582,970 bp. Its synthesis required the use of vectors that are often associated with large segments of genomic DNA, such as bacterial and yeast artificial chromosomes, known as BACs and YACs, respectively [30]. Segments of the M. genitalium genome were progressively combined by in vitro recombination in a BAC vector. When DNA instability in E. coli became a problem at the 290-kb size (half the M. genitalium genome), the authors changed their cloning host to S. cerevisiae, which can stably maintain longer sequences as artificial chromosomes. The 1/4 genome clones that could not be joined stably in E. coli were joined using transformation-associated recombination (TAR) cloning [31], in which each assembly was mixed with the TAR vector and transformed into yeast. Six DNA fragments were transformed at the same time, including three 1/4 genome segments, two fragments of one of the 1/4 genome segments and the TAR vector. Overlaps between the transformed DNA allowed for homologous recombination, which joined the fragments into a circular DNA species containing the Mycoplasma genome fused to the vector sequence. Based on Southern blots and PCR confirmations, at least 17 of the 94 clones tested seemed to contain the complete genome. One was shotgun sequenced to 7× coverage and matched the designed sequence exactly. Before sequencing, a NotI restriction digest was used to release the genome from the vector DNA. This band migrated at the appropriate 583-kb size.

Software enabling gene synthesis

The complexity of gene synthesis processes would not be possible without software applications that have specifically been developed for that purpose. Before DNA fabrication, the sequence is often designed de novo or redesigned from a pre-existing sequence. The design step could include combining different genetic parts in gene expression cassettes (Figure 4). DNA design also frequently focuses on codon usage, RNA secondary structures and other patterns in the sequence that might impact the levels of expression.

DNAWORKS (http://mcl1.ncifcrf.gov/dnaworks/dnaworks2.html) [32] and GEMS (http://software.kosan.com/GeMS) [33] are two software packages that provide users with a set of oligos with organism-optimized codon use, homogenous melting temperatures and minimal hairpin tendency. The GEMS package also permits inclusion or exclusion of specific restriction sites. Oligo Cuts (http://gcat.davidson.edu/IGEM06/oligo.html) is a simple web application for designing oligos for gene synthesis.

GeneDesign (http://slam.bs.jhmi.edu/gd/) [23] is a web application composed of multiple modules and is the most comprehensive software suite for gene synthesis. It includes modules for back- translating amino acid sequences using different codon usage tables or recoding a gene to preserve codon bias but yielding a very different DNA sequence. Silent site insertion and removal modules can be used to either insert or remove restriction sites, respectively, with synonymous changes to coding regions. For longer pieces of DNA, this software provides a top-down synthetic route by dividing long sequences (>5 kb) into shorter building blocks of ~600 bp that can readily be sequence verified and have junctions compatible with pre-specified joining methods. For ease of assembly, these building blocks are in turn divided into overlapping oligos with similar physical characteristics. Oligos have a target length of 60 nt and should exhibit similar melting temperatures.

**Figure 4.** Parts assembly using BioBrick™ parts. In this figure, BioBricks containing parts ‘A’ and ‘B’ are joined to illustrate the assembly process. BioBrick parts are flanked by a prefix sequence (green) containing restriction sites for EcoRI and XbaI, and a suffix sequence (red) containing restriction sites for SpeI and PstI. Each BioBrick is digested with an enzyme pair. The recipient, which will retain the plasmid backbone, is digested with two enzymes that cleave in the suffix sequence (PstI and SpeI) and the donor releases its insert through digestion with an enzyme each in the prefix and suffix (XbaI and SpeI). This restriction fragment is ligated into the recipient plasmid between the SpeI and PstI sites. Because cleavage with XbaI and SpeI leaves compatible, cohesive ends, their restriction products can be joined leaving a non-cleavable scar at the junction. The inserts between these flanking sequences, therefore, should not contain any of these restriction sites. Assembly of two BioBrick parts is directional and the product is itself a BioBrick, which enables multiple iterations of the standard assembly reaction to join additional BioBricks.
Box 3. Outstanding questions

Will it be possible to make error-free oligonucleotide synthesis?
The high error rate of oligonucleotide synthesis is the most crucial factor in DNA synthesis protocols today. Oligos without truncations and lacking point mutations would make DNA synthesis much faster and cheaper.

Will it be possible to miniaturize oligonucleotides synthesis?
Avoiding waste of excess oligonucleotides would save on expensive reagents and would, therefore, reduce the cost of DNA synthesis. This technology has been reported and its application for gene synthesis has been verified [38,39,59]; however, this synthesis scale is not available to mainstream consumers of oligos. Multiplexing oligo synthesis to synthesize in parallel all necessary oligos for a DNA synthesis project would also greatly improve productivity and reduce cost.

What will be the transformative applications of DNA synthesis?
DNA synthesis is still predominantly used as an alternative to cloning of natural DNA sequences. The attenuation of a virus by systematically changing the codons of its genes prefurges a new generation of biotechnology products that could not be developed from natural sequences. It is also possible that DNA fabrication will lead to abiotic applications used in computing, detections or smart materials that will have very little in common with traditional biotechnology products.

How can we automate the design of synthetic DNA molecules?
If DNA synthesis is still limiting, one can argue that our ability to design synthetic DNA molecules might already be limiting our ability to imagine new applications of the current DNA fabrication capability. Developing dependable models of the phenotype coded by a synthetic DNA molecule seems to be crucial to fully leverage this technology.

Will synthetic attenuated viruses be a viable avenue for vaccine development?
With their small genomes, viruses might be ideal models for complete genome redesign and synthesis. If synthetic viruses can be made with attenuated virulence that can still elicit a protective immune response, it would be possible to use them as vaccines, much like the current oral polio vaccine is used today. This has been successfully tested for the polio virus in mouse [20]. It still remains to be seen if tests in mice will be borne out in human subjects and whether this approach will be broadly effective for other viruses.

Box 4. DNA synthesis market survey

Blue Heron Biotechnologies, Inc. was incorporated in 1999 to provide commercial gene synthesis services. Since then, the market has been estimated to grow at an annual rate of 50–100%. It is likely to have crossed the US$100 million mark in 2007. It is currently dominated by Gene Art AG (http://geneart.com), DNA2.0, Inc. (http://www.dna20.com/) and Blue Heron (http://www.blueheronbio.com/). The rapid growth of the market attracted venture-backed startups such as Codon Devices, Inc. who, after deploying an aggressive commercial strategy for a few years, finally withdrew from the gene synthesis service market to focus on their synthetic biology activity. Several other companies who already provided molecular biology services have recently added gene synthesis to their offering. Companies well established in the oligonucleotide synthesis market, such as Integrated DNA Technologies (IDT), have also progressively added gene synthesis to their services. There is a fierce price war resulting in list prices for DNA synthesis that have dropped from approximately $1 per bp to $0.50 per bp over the last 12 months. Even though vendors often advertise a price per base pair, most of them quote each gene synthesis project individually. More often than not, the quote will include additional charges based on the sequence length and complexity, which make a substantial impact on the final price. Many gene synthesis projects require the synthesis of multiple constructs related to one another. Most companies refer to these constructs as ‘variants’ and the cost of synthesizing a variant is often quoted as a marginal fee that is much cheaper than the synthesis of the first construct in the series. Few companies, however, clearly spell out the criteria used to decide if a construct is a variant of another construct. Furthermore, the customer must usually specify explicitly the existence of re-usable sequences in an order. To explore to what extent gene synthesis companies could take advantage of the reuse of sequences in large gene synthesis projects, we requested quotes for a project including ~20 different constructs using a small number of genetic parts. Even though the 20 companies we contacted advertized comparable rates per base, there was approximately a twofold difference between the cheapest and most expensive quotes we received. Furthermore, the delay for fulfilling orders is not always predictable, especially on complex projects. Some companies have been unable to deliver some orders they accepted, whereas others refuse to quote the synthesis of some sequences.

Articles reviewing design issues associated with the combination of parts [34] or codon usage [27] have been recently published and software applications specialized in addressing these high-level design issues have been described elsewhere [23,24,35].

Technical challenges

It is fair to say that the major limitation to high-fidelity and low-cost DNA fabrication remains the synthesis of oligos (see Box 3 for additional outstanding questions). Because this chemical process generates many truncated and mutated oligos [36], DNA fabrication strategies need to mitigate the risk of uncontrolled mutations by verifying sequences during the construction process. Another limitation of oligo synthesis is the lack of a miniaturized process. The smallest quantities that are supplied to customers still contain far more material than is needed, thereby unnecessarily increasing fabrication costs. Radically new approaches to oligonucleotide synthesis could have a major impact on gene synthesis by reducing their cost or by increasing their quality. A variety of strategies, including parallel synthesis of oligonucleotides using methodology developed for the construction of microarrays [37,38] and microfluidic systems [39], have been described and are in active development.

Additionally, particular aspects of a target sequence can cause difficulties during fabrication. The presence of high degrees of secondary structure or highly biased AT or GC content can interfere with the annealing and assembly of single-stranded DNA, resulting in failure during LCR or PCA reactions. These properties of a sequence can usually be eliminated by juggling codon usage in a sequence. Additionally, the use of homology-directed assembly reactions on target DNA with repetitions of a DNA sequence often result in large deletions. When such repetitions exist, it is usually necessary to employ restriction-enzyme-based assembly strategies that avoid the need for sequence-specific annealing reactions. Finally, the synthesis of toxic genes can interfere with fabrication, because cellular intermediates are a component of all DNA fabrication strategies. Often the presence of genes can be predicted a priori. When known, the gene can be synthesized with a mutation, such as omission of its start codon or the intentional introduction of a frameshift mutation, to disrupt its activity. After fabrication, the mutation can be repaired by site-directed mutagenesis.
Box 5. DNA synthesis in the classroom

Several innovative education programs are fully embracing DNA fabrication. The International Genetic Engineered Machine competition (iGEM) has received abundant media attention since it began in 2003 [56,57]. Teams of mostly undergraduate students compete in the event. The goal of the competition is to design and assemble the most creative genetic systems by combining existing BioBrick parts and creating new ones. The competition climax is the convergence of all teams in Cambridge, MA, for the iGEM Jamboree. If iGEM is a summer project for most teams, some universities are taking advantage of this event to create innovative educational programs. Davidson College, for instance, uses iGEM to expose undergraduates to complex research questions at the interface of mathematics, computer science and biology [58]. This innovative approach to teaching, which combines lectures in the theoretical foundations of biology and mathematics with intensive laboratory work, has recently been recognized by the National Science Foundation through a multi-year grant to develop the program.

The Build-A-Genome (BAG) class at Johns Hopkins University is another innovative educational program. The BAG class is part of the Synthetic Yeast Genome project (http://baderlab.bme.jhu.edu/syntheticyeast/wiki/index.php/Sc2.0). Instead of relying on external vendors to synthesize DNA fragments, most of the synthesis is performed by undergraduates enrolled in the BAG class. The class starts with a two-week long molecular biology boot camp during which students are taught how to pipette, run a gel, setup a PCR, perform cloning operations and sequence-verify the resulting DNA. Upon completion of the initial training, student teams aim to synthesize 10-kb segments of the genome during the semester. Advanced students contribute new software resources and improved instrumentation and protocols. The software infrastructure uses Moodle open-source courseware and the course has been designed for portability to other schools. Students who have successfully completed a BAG semester have 24/7 access to the teaching laboratories for student-initiated iGEM projects.

These new student opportunities illustrate an important change in the way biotechnology can be taught. Early on, students majoring in physical and life sciences learn to work together in transdisciplinary teams. Their first experience of working with DNA is ‘making it’ instead of recovering it from biological samples. As a result, they approach biotechnology as an engineering domain where things can be built from the ground up. They can think outside of the box and, for them, DNA synthesis is the limit.

Concluding remarks and future perspectives

One day, DNA might be fabricated using a purely in vitro process, which does not involve intermediate passages through a host organism such as yeast or E. coli. Today, however, DNA fabrication still involves sophisticated cloning techniques but, nevertheless, a transition period has already emerged. Academic and commercial operators experiment with complex processes that combine the assembly of chemically synthesized oligos with cloning steps in attempts to construct long DNA molecules. Several companies have rushed to and sometimes later walked away from the gene synthesis market (Box 4 surveys the DNA fabrication marketplace). However, this does not mean only companies with extremely specialized techniques are capable of gene synthesis. Depending on the context, it might make sense to outsource DNA fabrication to an external vendor but, in other cases, there might be value in performing it in-house. In fact, gene synthesis projects are approachable by undergraduate students enabled by straightforward protocols and training in a small set of molecular biology skills. Two examples of such educational programs with an extensive DNA synthesis component are provided in Box 5. In any case, it is important to understand that the fabrication of small DNA fragments (<1 kb) is often very straightforward, but the assembly of longer DNA molecules raises several inherent technical difficulties that need to be understood. DNA fabrication is not a black box that would involve radically different techniques to those commonly used in a molecular biology laboratory, nor does it require expensive equipment.

References

16 Aslanides, C. et al. (1994) Minimal length requirement of the single-stranded tails for ligation-independent cloning (LIC) of PCR products. PCR Methods Appl. 4, 172–177
SciTalk needs you!

SciTalk, founded in 2005 by Ann Lackie and Peter Normington, is a free resource for fiction-writers that helps them to discover the exciting ideas, images and topics of modern science. It aims to encourage writers – novelists, poets, playwrights – to meet and talk to UK-based scientists, and to visit their labs and work-places to see how, why and where science is carried out, and to discover that scientists are real people, not just clichés. In return, scientists discover that collaborating with writers can be challenging and usually fun (and that writers are real people too).

Browse www.scitalk.org.uk and consider whether you would like to become a contributor too, or contact enquiries@scitalk.org.uk for further information.