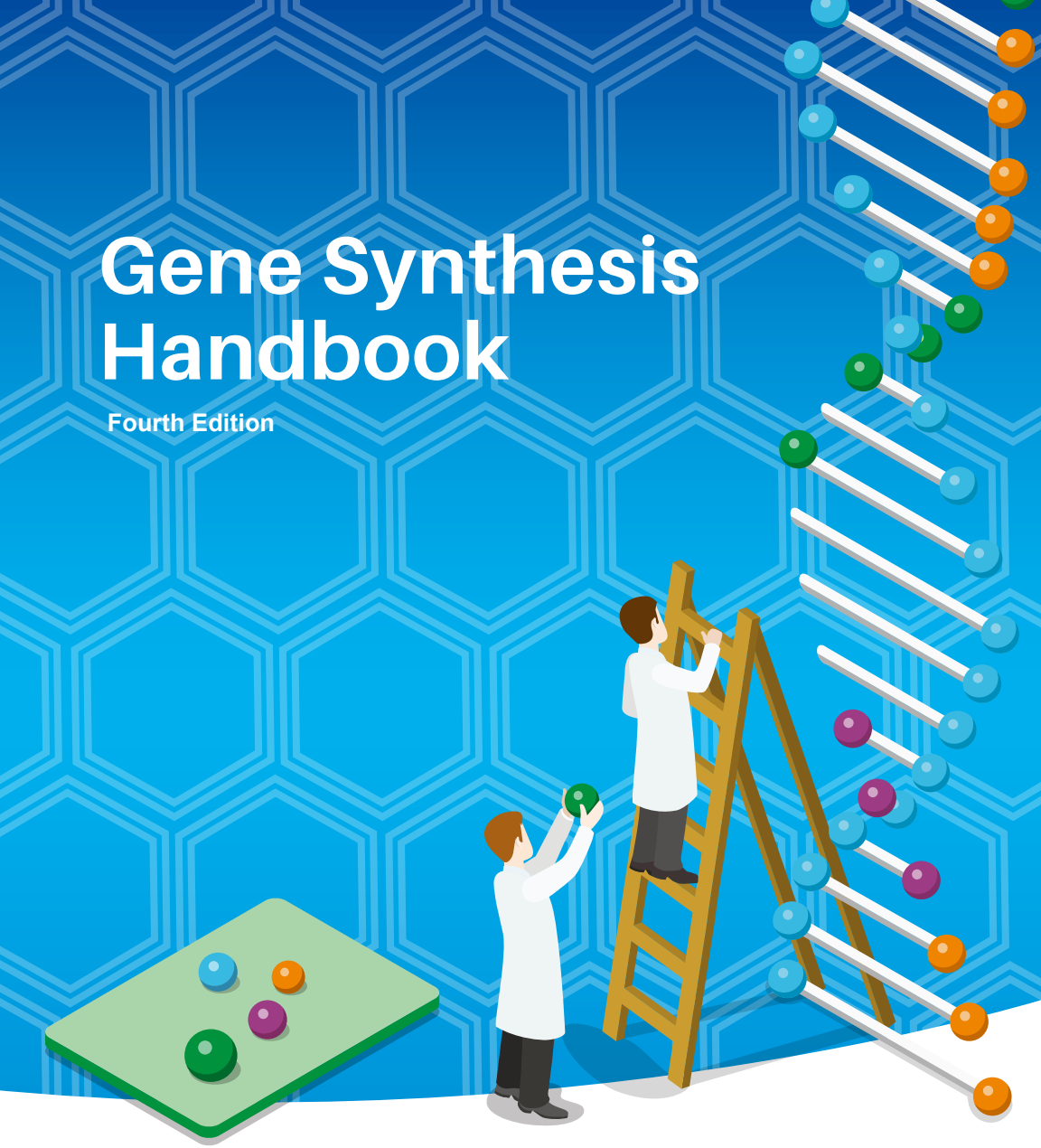
Gene Synthesis Handbook

Fourth Edition

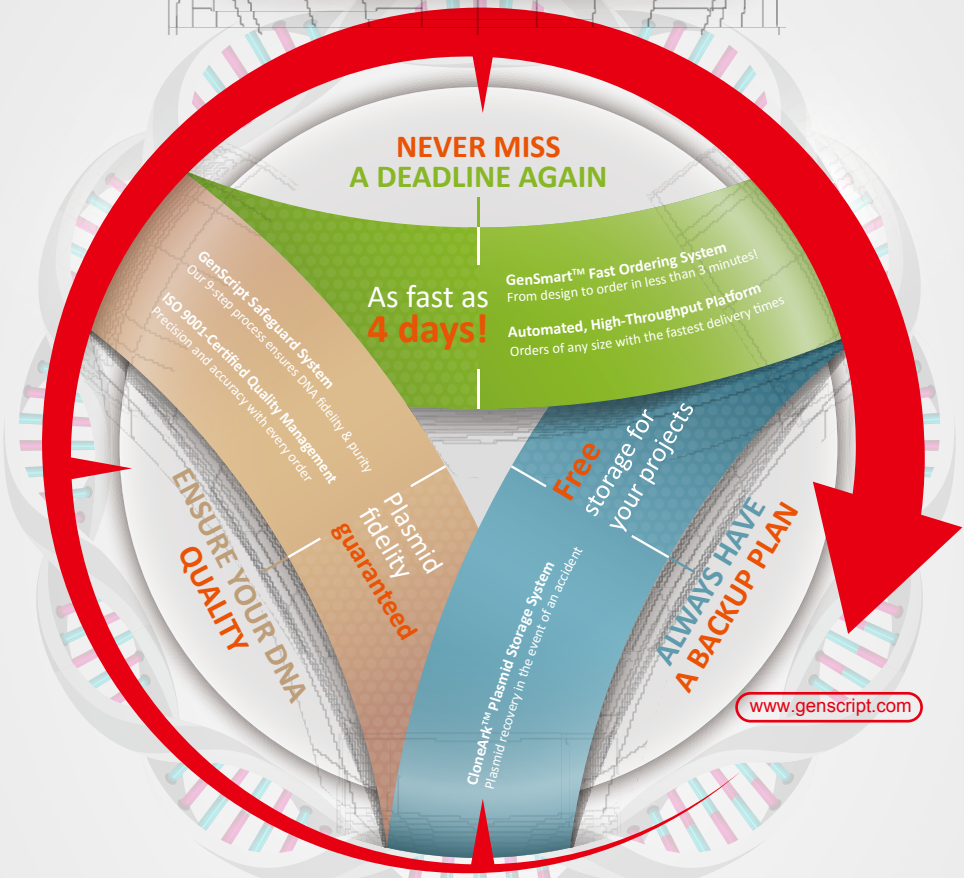


A Guide to Gene Synthesis Design and Applications

GenScript
Make Research Easy

Custom Gene Synthesis

Take Control of Your Research with ***GenScript***



Additional Features

PhD-level Support at Every Step

Gene Expertise from Over 600,000 Projects

Industry-leading OptimumGene Codon Optimization Technology

Table of Contents

Chapter One

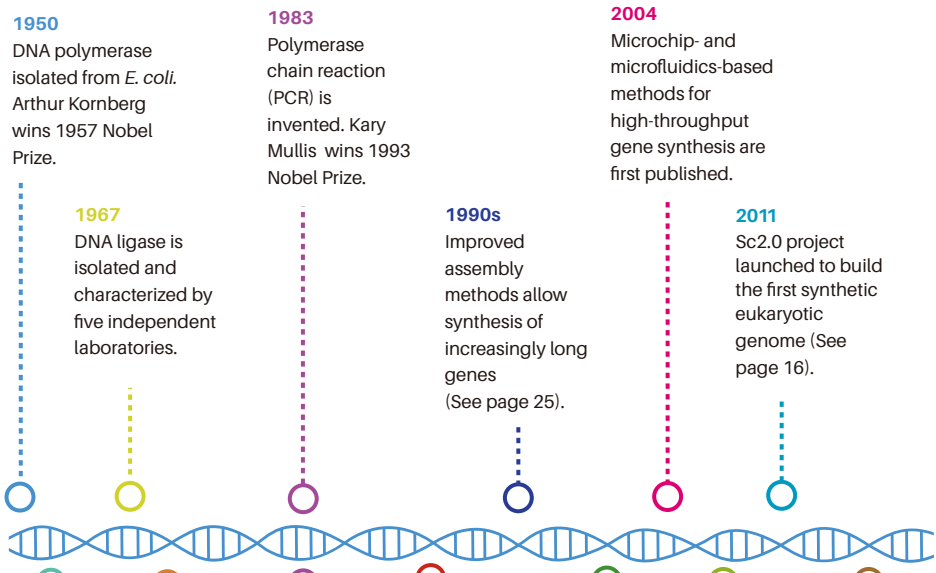
Introduction

Definition and History

Gene synthesis refers to the base-by-base process of chemically synthesizing a strand of DNA. Unlike traditional methods of DNA replication, gene synthesis does not require a template strand. Rather, gene synthesis involves the step-wise addition of nucleotides to a single-stranded nucleic acid molecule, which then serves as a template for the synthesis of the complementary strand. Sequences with varying lengths can be synthesized as stand-alone genes or building blocks corresponding to a larger gene, which are then enzymatically assembled together. The gene synthesis technology is a powerful and rapidly advancing tool, which has significantly affected all areas of biology with exciting and far-reaching implications.

The history of gene synthesis began in 1955, when Sir Alexander Todd published a chemical method for creating a phosphate link between two thymidine nucleotides; effectively describing the first artificial synthesis of a DNA molecule[1]. Based on this method, in 1970 Gobind Khorana's group reported the successful synthesis of an entire gene, the 77 kilo base (kb) pairs gene for a yeast tRNA, in the span of five years[2]. Development of phosphoramididite DNA synthesis by Caruthers and Matteucci in 1983 was fast followed by commercial availability of automated oligonucleotide synthesizer machines in 1989. Concentrated efforts in the 1990s, lead to the development of improved assembly methods that could enable the synthesis of longer genes. A decade later, the synthesis of an entire viral genome[3] in 2003 paved the way for the launch of the Synthetic Yeast Genome Project (Sc2.0) as the world's first synthetic eukaryotic genome project. This global collaborative effort aimed to create a novel, rationalized version of the genome of the yeast species *Saccharomyces cerevisiae,* comprised of 16 designer synthetic chromosomes encompassing 12 million bp of DNA. Scientists aim to build upon the success of the Sc2.0 project to tackle the challenges faced in synthetic biology, with the end goal of revolutionizing applications in cell biology and medicine.

Timeline of Gene Synthesis Technology Development



1950
DNA polymerase isolated from *E. coli*. Arthur Kornberg wins 1957 Nobel Prize.

1955
First DNA molecule synthesized when two thymidine nucleosides are joined by a phoshate link. Sir Alexander Todd wins 1957 Nobel Prize.

1967
DNA ligase is isolated and characterized by five independent laboratories.

1970
First synthesis of an entire gene, a 77 bp yeast tRNA, by Gobind Khorana's group.

1983
Polymerase chain reaction (PCR) is invented. Kary Mullis wins 1993 Nobel Prize.

1983
Caruthers and Matteucci invent phosphoramidite DNA synthesis.

1989
First automated oligonucleotide synthesizer machines become commercially available.

1990s
Improved assembly methods allow synthesis of increasingly long genes
(See page 25).

2003
First synthesis of an entire viral genome, that of phiX174 bacteriophage.

2004
Microchip- and microfluidics-based methods for high-throughput gene synthesis are first published.

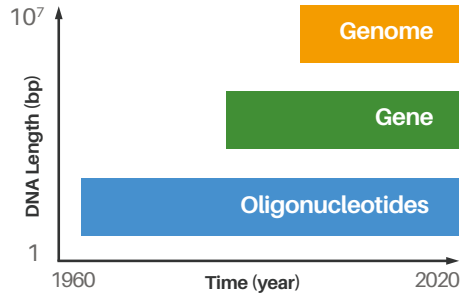
2008
First complete synthesis of a bacterial genome.

2011
Sc2.0 project launched to build the first synthetic eukaryotic genome (See page 16).

2014
Expanded applications beyond biomedical sciences, such as data science.

In the past 60 years, methods for *de novo* chemical synthesis of DNA have been extensively refined. Continual decreases in cost, advancements in assembly techniques and error reduction, and improvements in high throughput capabilities through automation and platform integration is rapidly making *de novo* gene synthesis the preferred method for molecular biology and bioengineering. These advancements along with improved methods for sequencing and amplification have now made it possible to synthesize genes larger than 1 kb in just a few days. Moreover, synthesis of entire genomes and cost-effective outsourcing to commercial providers is rapidly changing how biological research is being conducted[4]. Now, questions about the feasibility of biological research have shifted from technological limitations towards envisioning new paradigms for unravelling the mysteries of life and utilizing them for societal benefits.

Evolution of *de novo*-Synthesized DNA Over Time



* Adapted from Carr *et al.* 2009

Types of Synthetic Products

Gene synthesis technology can be used to generate a wide variety of nucleic acid products for diverse research applications. These products range from natural sequences to completely *de novo* nucleic acids in the form of DNA, RNA, or chimeric DNA-RNA backbones. The length of a synthetic nucleic acid can range from a few base pairs to several kb if assembled in a standard molecular biology laboratory, or over 10 kb if synthesized by specialized gene synthesis providers.

Dexoyribonucleic Acid (DNA)-Based Product

- **DNA Oligonucleotides,** synthesized as single stranded, short (18-22 nt) fragments, are commonly used as primers for sequence amplification, sequencing, and mutagenesis. They also can serve as probes in hybridization experiments. Oligo nucleotide libraries or pools can also be used for screening purposes.
- **Complementary DNA (cDNA),** corresponding to the open reading frame (ORF) of a given gene, can be synthesized for a variety of gene expression studies. Synthesized genes can encode either naturally occurring transcripts or custom-designed variants, including point mutations and reporter tags. In addition, synthetic cDNAs can also be modified for improved gene expression through codon optimization.
- **Genomic DNA** can be synthesized for the study of gene regulatory elements, mRNA processing, or to create artificial genes and genomes. Synthetic genomic DNA can be studied *in vitro* through the creation of transgenic animal lines or transfection of cell lines in culture. Conventional gene targeting of embryonic stem cells with disruption cassettes inserted through restriction enzymes[5] or Cre-loxP system[6] can be made simpler, more rapidly, and less expensively by using synthetic DNA in combination with newly developed methods for genome editing such as CRISPR/-Cas-9 nucleases, Zinc-Finger Nucleases (ZFNs) or Transcription Activator-Like Nucleases (TALENs)[7].
- **Extracellular DNA (eDNA)** is a component of the organic matter pool in the environment and play important roles in innate immune responses[8], thrombosis[9], cancer metastasis[10], and biofilm formation[11] among other biological functions. Gene Synthesis technology can fabricate any DNA sequence to further elucidate its role outside the cell.

Ribonucleic Acid (RNA)-Based Products

- **RNA Oligonucleotides** are used as primers for hybridization experiments and as components in CRISPR genome editing. Oligo libraries or pools of oligonucleotides can also be used for targeting and editing studies.
- **Interfering RNA** can be synthesized in a variety of formats to suppress gene expression. Small interfering RNA (siRNA) sequences of 21-25 bp can block gene expression by targeting complementary sequences within mRNA for degradation or sequestration. siRNAs synthesized as RNA oligonucleotides can anneal to form RNA duplexes, but these molecules are short-lived and often difficult to deliver into cells. The alternative is vector-based DNA constructs that encode short hairpin RNA

(shRNA) molecules which are easier to deliver through viral transduction, plasmid transfection or electroporation. This will allow the stable expression of shRNA under the control of inducible or tissue-specific promoters. Upon transcription, shRNA sequences form a stem-and-loop structure that is cleaved to produce the biologically active siRNA. The Targeted silencing of gene expression through siRNAs enables the study of genes whose complete knock-out would be otherwise lethal. Vector-based siRNAs delivered through lentiviruses can also be used cost-effectively to generate stable cell lines or transgenic animals[12].

Significance

Gene synthesis technology has revolutionized biology research and is starting to impact areas outside of life sciences. The scope of the applications that can utilize gene synthesis has expanded beyond traditional laboratory settings, even reaching disciplines like data science. Major advantages of gene synthesis technology can be described in two categories: overcoming limitations of previous technologies to simplify the creation of nucleic acid-based tools, and offering novel opportunities for creating the foundation of the field of synthetic biology and expanding its applications.

Simplifying Tools Development

Groundbreaking discoveries underlying the molecular cloning technology in the past several decades have evolved this valuable technology into one of the most powerful resources in science. However, despite recent improvements in recombinant tools, synthetically obtaining a clone of interest is hardly a fool-proof endeavor. Molecular cloning strategies have intrinsic limitations, which can not only reduce the pace and increase the cost of scientific discoveries, but also hamper the expansion of science as a whole. This is where gene synthesis can help complement and overcome the shortcomings of molecular cloning.

Gene Synthesis vs Molecular Cloning

Technique	Feature
Traditional molecular cloning	• Limited to sequences that appear in nature • Codon optimization is not feasible • Time- and resource-intensive for end-user
Gene synthesis of DNA fragments	• Customizable sequence; no template required • Error-prone; yields mixed pools of variants • Time consuming to verify synthesized sequences • Lower cost for high-throughput screens
Gene synthesis with subcloning into a vector	• Customizable sequence; no template required • Uniformly accurate sequences from a clone • Easy to verify synthesized sequences • Fast and cost-effective

The following summarizes major issues with molecular cloning strategies:

- **RNA yield and stability:** Starting material often needs to be extracted from a living organism. The unstable nature of RNA, inefficiency in extraction methods, and low mRNA copy number can significantly affect the efficiency of this initial step. A project risks failure if the starting material cannot be obtained.
- **Polymerase Chain Reaction (PCR) program optimization:** Robust PCR is key in both sequence amplification and PCR-mediated site-directed mutagenesis. Because of the unique sequence characteristics of each gene and potential complexities that may arise from site-directed mutagenesis, PCR optimization can be both time-consuming and expensive.
- **DNA yield:** In the process of cloning, DNA purification is required at several steps. At each step, the percentage of DNA recovery can vary from as high as 95% to as low as 60%. Low recovery rates and the presence of contaminants from the purification procedure can result in the loss of DNA.
- **Ligation efficiency:** The ligation reaction joining the gene insert and vector is a key step in the cloning process. Several factors, such as the insert to vector ratio, salt concentration, phosphorylation or de-phosphorylation steps, and solvent contamination can reduce the efficiency of the cloning process.

- **Mutations:** The multi-step process of molecular cloning, specifically insert and vector amplification, increases the chances of introducing unwanted mutations. While the use of proof-reading polymerases can reduce this possibility, their specific reaction conditions, such as sustained high temperature, can in turn lead to DNA damage and mutations.
- **Protein quality and yield:** The goal of molecular cloning is to obtain a high quality, functional protein at the desired amount. However, two factors can significantly affect this outcome. First, toxic genes introduced during the vector propagation step can increase the survival of mutation-containing transformants while leading to the death of correct transformants. This can ultimately affect the expression of a less potent version of the target protein. Secondly, constructs designed for heterologous expression often result in low protein yields due to several reasons, such as variations in codon usage across species. These factors can ultimately undermine the main objective of the cloning.
- **Engineered sequences:** The requirement for a naturally-occurring template limits the application of traditional molecular cloning where novel or customizable sequences need to be engineered. Although modifying sequences can be accomplished using a cloning strategy, complexities and time required, significantly hinders its application for novel scientific endeavors.

The intrinsic characteristics of gene synthesis technology have made it the preferred alternative for efficiently developing a variety of DNA tools. These features include:

- **Template-independent:** With this technology, any existing or imagined sequence regardless of its complexity or novelty can be synthesized *de novo*.
- **Limited-steps process:** Compared to steps involved in molecular cloning, which are numerous and prone to complications and errors, the steps involved in gene synthesis are limited and straightforward. Synthesizing genes from design to execution takes between one to three steps, depending on the length of the sequence.

- **Error-minimized:** The fully automated nature of gene synthesis significantly reduces the possibility of unwanted errors.
- **Customizable:** Whether used as a stand-alone technology or in combination with molecular cloning, gene synthesis offers endless options for customization. From the gene sequence and associated modifications to adaptability to diverse vectors, the synthesis process is highly amenable to customization. Codon optimization is one example of such capacity which can significantly increase protein expression in heterologous systems.

Enabling & Expanding Horizons of Synthetic Biology

The *de novo* synthesis of any nucleic acid sequence through gene synthesis enables the creation of a variety of biomolecules. Partial or complete enzymes, genetic elements and circuits, and metabolic pathways can be synthesized as core biomolecules for cellular assembly and incorporation. This significant feature of the gene synthesis technology provides an enormous capacity for the development of foundational compartments and tools for the field of synthetic biology.

As an emerging and interdisciplinary field, synthetic biology seeks to make the design, construction and optimization of biological systems easier, more predictable, and reliable. The ultimate goal of synthetic biology is to enhance possible biological functions for research applications and to develop therapeutics, materials, new means for information processing and data storage, as well as new sources of food and energy[13, 14]. This is achieved through the assembly of core biomolecules as engineered parts towards new or improved biological functions, with the ultimate goal of helping to improve living conditions and preserving the environment.

Knowledge and technologies from a wide range of disciplines, including engineering, biology, computer science, and mathematics have contributed to the development of this interdisciplinary field. However, among them, gene synthesis is considered the most enabling and makes up the foundation of synthetic biology. Information and tools generated by this technology have made the transfer of engineering principles into the biological discipline possible. Multi-part DNA constructs and biomolecular modules used in DNA regulatory elements, gene and protein networks, obtained through the use of gene synthesis, are essential components of engineering-based circuits or network

designs in biological systems for modulating gene activity and cell behavior. With extensive and direct applications in biomedicine, and implications or indirect impacts on other scientific fields, gene synthesis technology is considered a revolutionary tool in biology research in specific, and in sciences in general.

GenScript's comprehensive **Molecular Biology Services**, includes gene synthesis, mutagenesis, cloning, and plasmid preparation; enabling you to confidently move your synthetic biology ideas from design to application. Synthesize innovation today with GenScript!

Chapter Two

Applications

The revolutionary technology of gene synthesis has extensive and diverse applications in basic research and applied sciences. Gene synthesis is a major foundation of the current revolution in the biomedical sciences and biotechnology, impacting everything from healthcare to conservation. Beyond life sciences, gene synthesis is enabling data and material sciences to test theoretical hypotheses, conceive new technologies or applications, and even develop new areas of science, such as DNA nanotechnology. Ongoing improvements in synthesis fidelity and efficiency, process automation, cost reduction and design algorithms will further expand the application of this technology.

Core Applications

The ability to design and synthesize custom nucleic acid sequences allows for three unique core applications. These core applications can be used in several ways: directly in an experimental setup/procedure, as the basis/precursor for the generation of other tools and reagents, or as a module in a multi-component tool or application.

1 Generating Nucleic Acid-Based Reagents

DNA- and RNA-based synthetic products of gene synthesis are used as essential tools and reagents in biotechnology and biomedical research. These products range from custom oligonucleotides, plasmids, genomic and eDNA, cDNAs and ORF clones as well as interfering RNAs. The extent of the application of these reagents and tools in all areas of life sciences is so vast and fundamental that without them none of the current scientific advancements in disease pathogenesis, drug discovery and understanding life at the molecular level would have been possible. Here are a couple of ways these tools are used in research today:

- **Oligonucleotides:** As the most-widely used reagent, oligonucleotides are used as hybridization probes in bioassays, primers for nucleic acid amplification, gene assembly, sequencing, or for introducing mutations. Modifying these products with various tags, such as fluorophores, further expands their application in detection or quantification assays, such as quantitative real-time PCR. In recent years, advancements in high throughput synthesis technologies have further expanded the range of applications for this valuable tool, offering hundreds of custom oligonucleotides in library or array formats in single or pool versions for screening purposes.
- **cDNAs and ORFs:** Synthesis of cDNAs and generation of ORF clones have helped scientists overcome the shortcomings of molecular cloning and expand its utilization in science. Traditionally, an ORF clone is considered a plasmid that contains a

protein coding DNA insert which upon transfection into the right host, instructs the mRNA translation machinery to express the full length, functional protein. Synthesis of an ORF skips the success-limiting step of obtaining the full-length gene sequence through PCR and restriction enzymes in conventional cloning. Moreover, one can easily assemble several gene cassettes to form fusion proteins while ensuring the maintenance of the correct open reading frame. Synthesized cDNAs and ORFs can be delivered bare or cloned into any type of vector. While bare ORFs are less costly and more suitable for quick high-throughput screenings as well as *in vitro* transcription and translation, they are less advantageous when compared to ORF clones. Insertion of an ORF in a vector (a) protects the synthetic DNA from degradation, (b) facilitates clonal amplification in a transformation-competent host, and (c) simplifies shuttling of the synthetic gene into a variety of vectors for transient of stable expression.

Synthetic short oligonucleotides and genes can be used as reagents for the synthesis of vectors, eDNA or the entire genome of an organism. For short eDNAs, a one-step synthesis will generate a reagent for direct application in an experimental setup. For large eDNAs or genomes, ORFs and regulatory elements are first synthesized and later, with the use of enzymes or homologues recombination, assembled to create the whole genome. Introducing these tools and reagents as functional units or modules of a network into various biological hosts, allows the study of regulatory elements, transcription and translation machinery, function of DNA sequences outside the nucleus, and generation of new life forms.

- **RNA interfering molecules:** RNA interference (RNAi) was originally named "post-transcriptional gene silencing" and "quelling" based on the initial discovery that RNA inhibited protein expression in fungi and plants. The new terminology was later coined in late 1990s where a similar phenomenon was discovered in *C. elegans*[15]. Further work on elucidating the underlying mechanism of RNAi combined with the power of gene synthesis, provided a plethora of reagents for studying mammalian gene expression and developing therapeutics. RNAi reagents, such as siRNAs, shRNAs, and miRNAs, are the preferred tools for loss of function studies, target validation, and gene expression analysis in signaling pathways induced by natural ligands or toxic compounds. RNAi molecules can be used directly or through vectors in transfection assays *in vitro* and *in vivo*. Compared to dominant negative constructs, ribozymes, or knock-out animal strategies, the RNAi approach is less expensive, more straightforward, efficient, and target-specific. Advancements in design algorithms and modification technologies is leading to the development of the second generation of RNAi tools with increased stability, greater specificity, and with less chance of invoking an immune interferon response.

By synthesizing and cloning Rhinovirus RV1BV4 capsid protein and human N-myrisotyl transferase, GenScript helped Bentley et al. demonstrate how the interaction between these two proteins can induce an inflammatory immune response in the respiratory system.

Recombinant Myristoylated Rhinovirus VP4 Protein Activates Pro-Inflammatory Gene Expression in HEK293 Cells and Primary Alveolar Macrophages. Am J Respir Crit Care Med, 2017.

2 Codon Optimization

Proteins are effector molecules in cellular machinery and as such, studying or modulating living organisms requires access to functional proteins in large quantities. Molecular cloning allows heterologous gene expression in which the protein coding sequence from one organism can be transferred into another host or model organism for convenient expression and extraction of proteins in a large scale. Heterologous expression enables the production of pharmaceuticals of human or animal origin without the health concerns associated with production from their native host. Unfortunately this strategy often suffers from lower than ideal rates of protein expression.

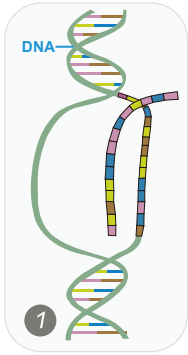
A common cause of low expression levels is codon bias or the variation in codon usage frequency among different species. Translation of a heterologously-expressed gene can stall due to tRNA scarcity in the host organism which in turn leads to low protein levels and increased rates of improper protein folding. While overexpressing rare tRNAs can help remedy this problem to some extent, it can interfere with translational machinery that regulates all endogenous protein turnover and neglect to address other optimizable features of the coding sequence. Taking advantage of codon degeneracy is the alternative and preferred approach: the DNA sequence can be changed in a way that does not alter the final amino acid sequence, while improving the efficiency of transcription, mRNA stability, translation, and protein folding.

The process through which an amino acid sequence is rendered into a DNA sequence with codon usage suitable for maximum expression in an organism is called "codon optimization". Codon optimization not only helps with codon usage bias, but also improves other features of a DNA sequence that affect the efficiency of transcription, proper splicing and processing of mRNA, and its stability. Optimizing codons can be achieved through the laborious and costly method of side-directed mutagenesis, but application of gene synthesis can make this process efficient, fast, and less expensive[16]. Development of a variety of algorithms enables the design of the most optimized sequence for increased expression, sometimes upwards of 100 folds when compared to the non-optimized versions of the gene.

Codon optimization is widely used in different disciplines. For example, in structural biology studies in basic science,

codon optimization is employed to ensure sufficient yield of soluble proteins. In translational research and drug development, codon optimization helps with enhancing the immunogenic response to DNA vaccines. Widespread application of codon optimization is also adopted in agriculture for a variety of purposes, such as improving crop traits to increase yield or confer pest resistance.

Factors Influencing Protein Expression

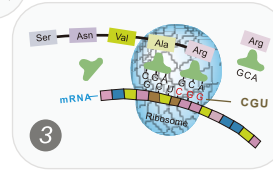


1. Transcription

- cis-regulatory elements (TATA box, termination signal, protein binding sites, etc.)
- chi sites
- polymerase slippage sites

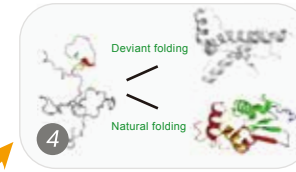
2. mRNA processing and stability

- cryptic splice sites
- mRNA secondary structure
- stable free energy of mRNA



3. Translation

- codon usage bias
- ribosomal binding sites (e.g. IRES)
- premature polyA sites



4. Protein folding

- codon context
- codon-anticodon interaction
- translation pause sites

GenScript's **OptimumGene™** PSO-driven algorithm, patented in 2012, efficiently optimizes sequences by simultaneously taking into account 200 variables that are known to influence expression, resulting in up to 100-fold higher protein yield than the native sequence and up to 50-fold better results than other codon optimization design tools. Improve your expression levels today!

GenScript's Optimum™ gene optimization algorithm and gene synthesis services, allowed Ramezani *et al. to demonstrate that a sequence-optimized variant of the* Pertuzumab gene can improve its production as a biosimilar drug.

Improving Pertuzumab production by gene optimization and proper signal peptide selection. Protein Expression & Purification, 2017.

3 Mutagenesis

Modification of DNA through mutagenesis is a well-established laboratory technique. These mutations help with understanding the function of a protein or genetic pathway, or providing new proteins with novel or enhanced functions with potential for therapeutic and industrial applications.

Traditional methods for introducing mutations include subjecting live cells or animals to UV radiation or chemical mutagens. These methods, though useful in reverse genetic screenings, have several drawbacks: (a) they usually lead to high rates of lethality and low rates of useful mutant phenotypes, (b) they are random for nature and as such require laborious efforts to investigate their outcome and make correlations to the location or frequency of the mutations, and (c) mutations may be biased and as a result limit the range of possible mutants to only a fraction of those that are theoretically possible or that can be obtained through other strategies. An alternative approach is site-directed mutagenesis, which introduces variation at specific and desired locations within DNA. This strategy, which is based on having prior knowledge of DNA sequence(s), can be performed using PCR-based methods. Mutated primer sequences are first created and used for introducing mutations through the amplification process, and finally products are subjected to sequence verification. While this approach provides specificity and is less labor intensive, it still can be tedious, error-prone, and costly.

Generating mutations via gene synthesis overcomes the shortcomings of traditional methods and provides new capabilities. Using this strategy, small or large number of systematic or random mutations can be created in various parts of a gene and regulatory elements, with higher efficiency and savings in effort and time. This results in more experimental control over the types or number of mutations in any mutant strain or gene.

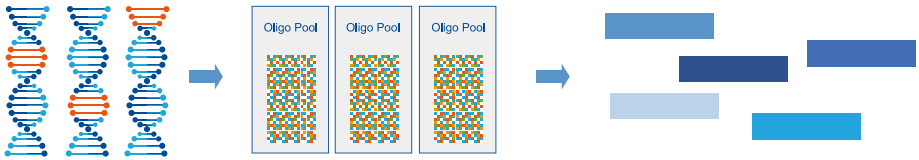
GenScript can deliver sequence-guaranteed mutants within two weeks. Save your lab the hassle of conducting site-directed mutagenesis, sequence verification, and error correction. Get your mutants on the timeline that best suits your needs with GenScript's **Express Mutagenesis Services.**

3.A. Gene Disruption Constructs

Targeted deletion of a gene, a series of genes, or regulatory elements of a gene is a useful tool in understanding biological processes. Gene synthesis helps with the generation of custom gene disruption constructs that can be used for *in vitro* studies or to generate mutant strains or knock-out animal models. Using this technology overcomes the limitations of conventional methods, such as transposons, restriction enzyme digestion and PCR-based methods[16]. Specifically, the lack of adequate restriction sites, non-specificity, and the requirement for large fragments of DNA can be a hindrance to generating multiple or even single mutations. Assembly of smaller synthetic, mutant gene compartments allows the construction of mutant gene cassettes for faster and more efficient investigation of genomic sections. Synthesis of a variety of ORF clones containing mutations or deletions that are expression-ready, facilitates the discovery of new protein functions, generation of proteins with improved or novel functions, and investigating disease pathogenesis. Application of gene synthesis results in increased accuracy when generating mutant constructs, as well as savings in both time and money during experimental execution.

3.B. Gene Variant Libraries

Gene synthesis facilitates the discovery or design of new protein variants (mutants) with properties desired for medical, industrial, and agricultural applications. Gene variant libraries allow for systematic, unbiased investigations, such as high-throughput screening for discovery biology, drug development, directed evolution for protein engineering, and *in vitro* molecular optimization to generate mutant proteins with improved or novel properties[14]. A synthetic DNA library may also be used to construct novel genetic circuits encoding enzymes that constitute biosynthetic pathways for metabolic engineering. They are also useful in identifying novel gene regulatory sequences or to identify minimal domains critical for protein structure and function. Gene synthesis allows the construction of a variety of custom libraries, such as site-directed mutagenesis, scanning point mutation, randomized and degenerated, truncation variant, precision mutant and combinatorial assembly libraries suitable for different experimental objectives. These libraries are fast becoming the preferred tool of choice for providing reliable, accurate, and reproducible results with time and cost efficiency.



GenScript's expertise in *de novo* gene synthesis, and its proprietary **GenPlus™** High Throughput Gene Synthesis Technology, allow for the synthesis of gene variant libraries of any size and complexity with unprecedented efficiency. No matter your DNA library construction goal, GenScript can provide a tailored service that meets your needs!

Applications of Core Products of Gene Synthesis

Type	Application
cDNA/ORF	• Over-expression or heterologous expression for functional, enzymatic or structural studies
Customized coding sequences	• Expressing fusion proteins • High-level protein expression from codon-optimized sequences for functional, enzymatic or structural studies
Promoter-reporter constructs	• Monitoring gene expression downstream of manipulations to transcription factors, signaling cascades, etc.
Genomic DNA	• Creating synthetic genes or genomes to obtain novel functions • Studying gene structure, regulation, and evolution
Mutant sequences	• Functional studies (promoter-bashing, amino acid substitutions) • Protein engineering (rational design or screening/directed evolution)
RNAi constructs	• Gene regulation studies Intercellular communication
Extracellular DNA	• Biofilm formation Intercellular signaling as in cancer metastasis
Gene variant libraries	• Promoter optimization • Functional characterization of any gene or protein • Optimizing protein structure and function • Improving stability and affinity of enzymes • Optimizing antibody affinity • Creating and screening a large number of variants when designing a metabolic pathway

Broad Applications

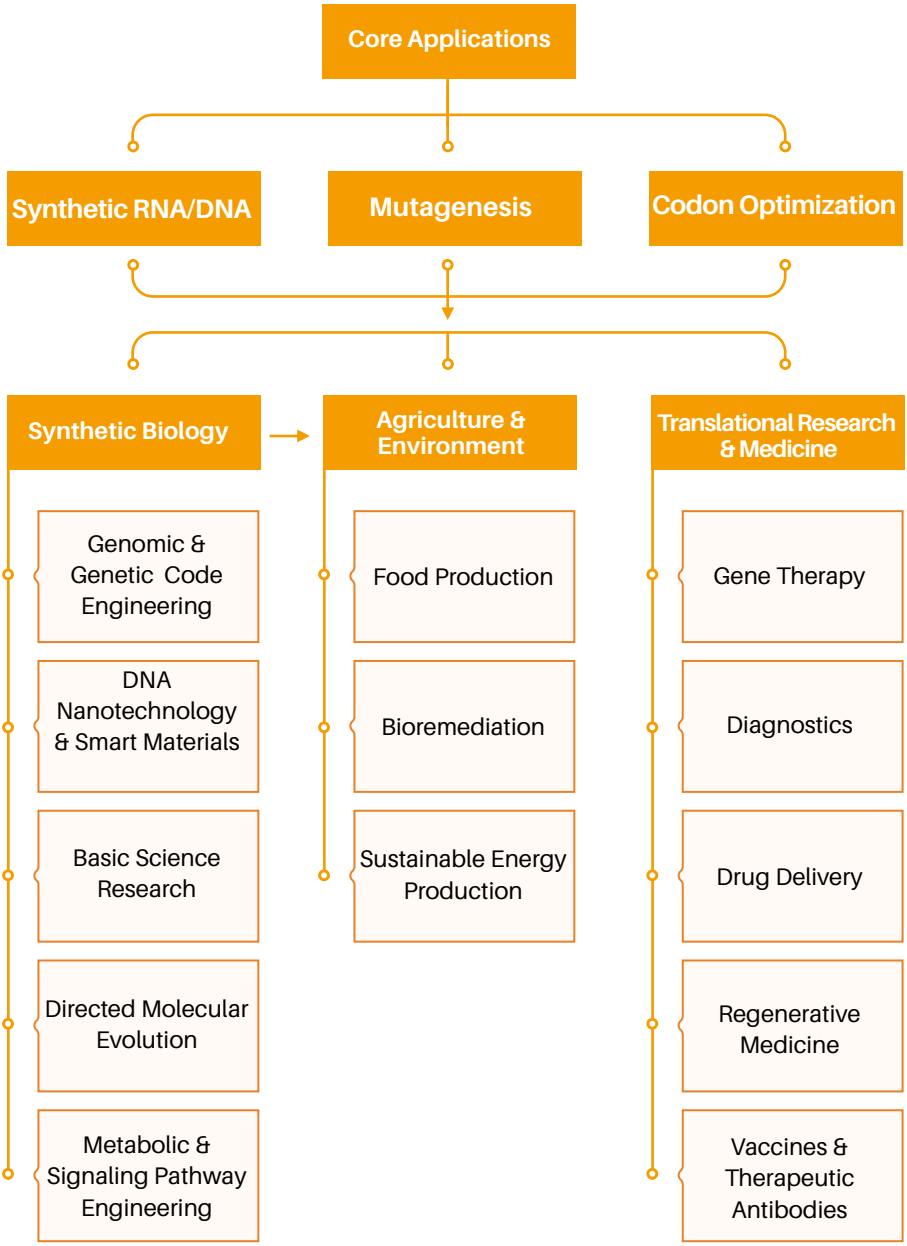
The core applications of the gene synthesis technology can be used for broader applications in basic life sciences, medicine, engineering, and agriculture. These applications not only provide novel solutions to existing and emerging problems facing society and the environment, but also create and define new scientific areas.

1 Synthetic Biology

About thirty years ago, the discovery of restriction enzymes and their application in constructing recombinant DNA molecules was predicted to lead us into the era of synthetic biology for constructing and evaluating new gene arrangements with unparalleled efficiency and accuracy[17]. Now, that prediction has turned into a vibrant and interdisciplinary field that seeks to make the design, construction, and optimization of biological systems easier, more efficient, reliable, and predictable. Using rational design strategies, synthetic biologists modify, combine, and repurpose existing natural genetic elements to create new parts and systems. Their approach is based on standardization of genetic modules, decoupling of complex biological problems into simpler units that can be solved or pieced with other units to create entirely new systems; reducing complexity and increasing efficiency of engineered biological systems[18]. The ultimate goal is to use these novel synthetics towards deepening our understanding of natural motifs and networks, developing new medicinal drugs and devices, new means of information processing and data storage, as well as sources of food and energy that can simultaneously help human health and preserve the environment.

Gene synthesis technology has enabled synthetic biology scientists to access high quality genetic material in large quantities, allowing for novel projects and applications to be pursued. Using gene synthesis to create genetic building blocks and modules for mixing and matching to build networks and circuits has fostered the development of novel methods and expanded the scope of gene synthesis applications. While early synthetic biology research focused on engineering networks in prokaryotic hosts, it is now the mammalian synthetic biology that is taking the most advantage of the powers of gene synthesis through manipulation of mammalian genomes and cloning large DNA circuits. Biomolecular tools generated by synthetic biology include tools to regulate RNA, transcription, protein turnover and signaling pathways. These tools, such as Biobricks™, are offered as standard interchange parts that can be assembled in a variety of combinations to form complex systems and devices for diverse applications.

Broad Applications of Gene Synthesis Products



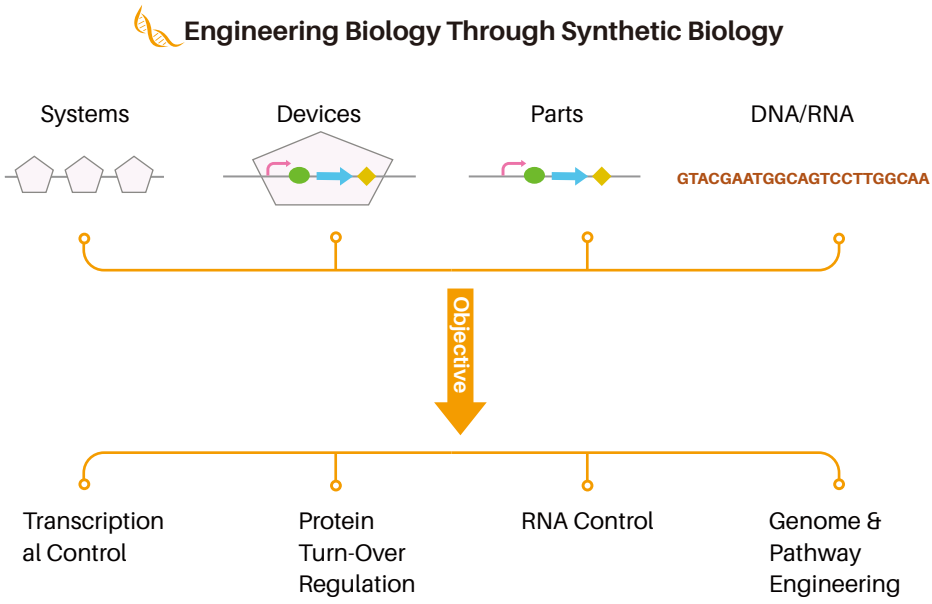
1.A. Metabolic & Signaling Pathway Engineering

The goal of metabolic engineering is to optimize microbial metabolic pathways to improve product yield. However, with the application of synthetic biology the scope of metabolic engineering has greatly expanded to not only increase yield in a more efficient and cost-effective manner, but also to produce a wider range of metabolites with unique features. Following the main components of biology's central dogma – transcription, translation and RNA processing – molecular control over metabolic and signaling pathways is first studied and quantified. This is then followed by the design and synthesis of either improved natural parts or completely novel elements, which are inspired by various natural components from various organisms or pathways. Assembly of these synthetic parts into circuits and networks can ultimately lead to engineered systems that can be incorporated into a new host to provide the desired metabolic output.

In addition to quantification studies on the rate-limiting steps and substrate concentration, synthetic biology approaches have also been focused on engineering the following aspects of protein expression for adjusting cellular metabolism[19]:

- **Promoters** for regular or constitutive activation through mutagenesis, recombination, or assembly of components taken from different organisms and pathways. Given that mutations often interact in a cooperative and non-linear manner with regard to metabolite production[20], synthetic biology has enabled multiple and simultaneous mutations in several genes involved in the transcriptional machinery in an approach currently called the Global Transcription Machinery Engineering (gTME).
- **Ribosome binding sites** (RBS) based on thermodynamic models of translation initiation which have led to the development of an RBS Calculator. Based on the desired translation initiation rate, gene sequence, and host organism, this algorithm predicts a customized RBS to achieve a targeted translation initiation rate[21].
- **RNA-focused parts** to modify RNA degradation affecting steady-state expression level. These parts include modular RNA elements, RNA regulators, and RNA aptamer domains. Moreover, RNA-related sequences can also be engineered. A prominent example is the Tunable Intergenic Regions (TIGR), which are synthetic RNA constructs that include two hairpins joined by an RNase cleavage site, and can be used to connect co-transcribed ORFs[22].
- **Spatial organization of cellular components** of a pathway to increase the local concentration of pathway intermediates and exclude competing cytosolic pathways. RNA and DNA scaffolds are examples of such engineering attempts[23, 24, 25].

Combinatorial application of these tools allows the generation of chimeric and novel pathways for optimized interaction in a heterologous host to achieve metabolic goals. In addition, synthetic biology approaches are used to engineer membrane receptors and protein domains involved in enzymatic activity, protein-protein interactions to develop networks for detecting unnatural small molecules, stimulating the immune system, and enhancing metabolic control[14]. The executive host of such approaches, which are kept in their natural state or engineered, can be isolated mammalian or plant cells as well as prokaryotes in single or ecosystem form. Engineering microbial ecosystems is currently being tested for a variety of purposes, such as bioremediation and therapeutic purposes[26].



GenScript is proud to be a contributing partner in the Sc2.0 International Consortium whose goal is to build a designer synthetic eukaryotic genome of Saccharomyces cerevisiae.

Design of a synthetic yeast genome. Science, 2017.

1.B. Genomic & Genetic Code Engineering

The precise manipulation of one or more genetic loci or amino acid codons for a specific outcome is the main objective of gene and genomic engineering. Defining the desired outcome is application specific and includes correcting deficiencies, imparting novel functions, or *de novo* synthesis of a new system or organism. Synthetic biology has helped to improve the specificity of such engineering objectives. Moreover, it has opened up new avenues in design and application, which would have been impossible to conceive and implement in the past.

Precise manipulation of mammalian genomes was first made possible by site-specific recombination. This method, based on bacterial or fungal recombinases (such as *Cre* recombinase) catalyzing the recombination of a pair of short target sequences (such as LoxP sites), can lack specificity, though. Using synthetic biology, conditional gene knockouts with increased specificity can be created. For example, an expression cassette containing two fragments of a split *Cre* recombinase from two different tissue-specific promoters has been synthesized. This allows the induction of a knock-out response to a combination of two sequences representing the same gene to achieve higher specificity[27, 28]. Moreover, synthesis of programmable DNA nucleases, such as TALENs and CRISPR, that can introduce site-specific DNA breaks leading to localized homologous recombination, has greatly improved the efficiency of genomic engineering. Furthermore, these methods offer the possibility of simultaneous allele mutations, although with this comes an increase chance of off-target cleavage. Recent advancements in Multiplexed Genome Engineering (MAGE) has been shown to reduce this chance. This oligonucleotide-mediated recombination approach allows for editing multiple genetic loci in parallel and is demonstrated to be useful in improving the yield of industrial product[29]. Similar attempts are under way in mammalian and plant systems for potential gene therapy and genetic improvements in crops, respectively[30]. Application of MAGE in microbial populations, which enables the generation of divergent and complementary phenotypes to obtain engineered ecosystems with desired functions, is also underway[26].

Engineering the genetic code is another emerging area whose conception was solely based on the gene synthesis and synthetic biology techniques. This area is comprised of (a) codon swapping through changing the genome sequence, such as synonymous codon swaps in one or more genes, (b) codon introduction through assigning a new amino acid to a codon without removing its original function, such as UAG usage as a stop codon while maintaining its coding for selenocysteine and pyrrolysine, (c) codon reassignment through changing the amino acid assignments of one or more codons across the whole genome, and (d) codon creation through adding a new codon to the translation code table, such as using codons with unnatural bases or quadruplet codons[31]. Engineered genetic codes can help with overcoming the challenges of the universal genetic code (such as viral infection and undesired horizontal gene transfer), understanding origins of life, and providing potential for new and useful protein functions with diverse applications.

Advances in gene synthesis and synthetic biology in fabricating long DNA sequences have also paved the way for current efforts towards *de novo* synthesis of whole genomes. With the experience gained through the successful construction of mouse mitochondrial genome[32] and a bacterial minimal genome[33], a consortium of scientists are currently working on synthesizing the modified version of yeast genome[34]. Such custom-made organelles and organisms are used to further understand the fundamentals of life and develop novel tools and products.

The International Genetically Engineered Machine **(iGEM)** competition is the premiere Synthetic Biology competition, with separate divisions for high school students, undergraduates, and researchers/entrepreneurs. Teams use standardized biological parts, both existing and new, to build novel biological systems and operate them in living cells, tackling real world problems facing society today. GenScript is proud to sponsor iGEM teams and to provide gene synthesis services that allow competitors to bring their designs to life!

1.C. Directed Molecular Evolution

Another area where synthetic biology has helped advance is artificial evolution. Directed molecular evolution aims to use a variety of synthetic biology-made parts and circuits to elucidate gene and protein function, improve protein and pathway activities, and fabricate novel proteins. From simple mutagenesis to shuffling receptor domains and manipulating ligand binding specificity, artificial evolution is providing a short-cut for obtaining enzymes, receptors, pathways and products. Development of high-throughput techniques, such as MAGE, are further advancing the screening of artificially-evolved molecules for therapeutic and industrial applications[14, 16].

1.D. DNA Nanotechnology & Smart Materials

One of the main problems in our times, currently being dubbed the information age, is the storing of data generated worldwide, which is predicted to reach 40 zigabytes (ZB) by the year 2020[35]. With a perspective of using smaller mediums rather than bigger data storage centers, the key to solving this issue can be found through the use of the DNA molecule. Billions of years of evolution have optimized DNA for transmission and storage of data in a way that is both efficient and reliable. Theoretically, all 40 ZB of man-made data can be stored in just 90 grams of DNA[36]. Initial attempts to use DNA as an archival medium could go only so far because of the difficulties faced when attempting to synthesize long stretches of DNA. However, with the current technological advancements in DNA synthesis and sequencing, synthetic biologists are fast making progress in storing large datasets in DNA. Although current limitations such as lack of random access read, rewritability features, and ability to decode data in the distant

future is keeping this novel solution at the experimental level, the multitude of advantages the DNA-based storage offer is foreseen to reshape the future of communication and data storage in the near future[36, 37].

Similarly, advances in synthetic biology has allowed the realization of an array of theoretical ideas in the field of DNA nanotechnology. Pioneering work of Seeman in the early 80's, which established the tile-based assembly method[38], combined with the origami method put forth by Rothemund at the turn of this century[39] have facilitated the fabrication of a variety of DNA nanostructures, such as nanosensors, nanoelectronic, nanophotonic devices and scaffolds. These programmable materials can be built in two or three dimensions and serve as a useful tool to construct structures with a variety of sizes and complexity with applications in biosecurity, product barcoding, protein structure studies, basic science research, diagnostics, and therapeutics[40].

Advantages of DNA-Based Data Storage

- Storage capacity 1,000,000-fold higher than current platforms
- Reproducible through fast and cost-effective PCR
- Stable and not subject to sequence change
- Passive and cost-efficient maintenance
- Hack-proof based on lack of remote access
- Long lifespan across thousands of years
- Devoid of technological obsolescence

1.E. Improving Traditional Research Methods

While basic science research was the birth place of synthetic biology, it has been the first field to benefit from its existence. One of the earliest applications of synthetic biology in basic science research was the use of transcription factor circuits in studying gene expression and regulation. The first generation of these circuits was comprised of a combination of natural and synthetic parts. For example, the *LacI* circuit, which is still widely used in laboratories, consists of a naturally existing bacterial transcription factor lac repressor (*LacI*) and a synthetic promoter containing the *LacI* transcription factor-binding sequence. This synthetic circuit can sense the expression of specific endogenous transcription factor as an input signal upon binding of the small molecule isopropyl-β-d-thiogalactopyranoside (IPTG). Alternatively, the promoter sequence of an endogenous gene can be fused upstream of a synthetic transcription factor coding sequence to drive its expression.

The second generation of transcription factor-based circuits, comprised of synthetic programmable transcription factors, is fast replacing the early types as a result of the high level of customization they offer. These include Zinc Finger-containing Transcription Factors (ZF-TF), Transcription Activator-Like Effectors (TALEs), and Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-based regulators. Customization of various elements of these systems has greatly expanded the repertoire and specificity of their application in basic science studies. Moreover, by fusing enzymes such as DNA methyltransferases that catalyze the addition or removal of chromatin marks, further studies on histone modification and ultimately gene regulation is made possible[41, 42, 43].

Synthetic gene circuits comprised of RNA and transcription factor elements are another synthetic biology-generated tool with basic science application. These circuits help with testing reverse engineering algorithms with positive and negative feedback loops to reveal the highly interconnected network of gene regulatory elements and to control cellular behavior[44, 45]. Moreover, rebuilding synthetic signaling networks and studying them in an exogenous context, such as expression in a heterologous host, simplifies the study of the complex and interconnected endogenous signaling pathways that may compete or affect a given network[14].

In addition, development of novel high-throughput methods through the synthesis of tandem repeats of short DNA sequences in libraries have expedited the discovery of strong transcription factor-binding motifs or gene and protein regulatory networks. For example, barcoding promoter variants with short DNA sequences and measuring their activity by high-throughput RNA sequencing, named "massively parallel reporter assay", has allowed the study of a large number of variants of enhancers and helped with the discovery of conserved regulatory motifs within enhancers[46].

Codon optimization, gene synthesis and cloning services of GenScript helped Zhang et al. to engineer the microalgae Synechocystis for generating an algal-based biofuel.

Construction and application of the Synechocystis sp. PCC6803-ftnA in microbial contamination control in a coupled cultivation and wastewater treatment. J of Environmental Sciences, 2016.

2 Translational Research and Medicine

A common goal among scientists in different disciplines is to develop novel approaches for disease diagnosis and treatment. Achieving this goal has been made more practical through the application of gene synthesis and its off-shoot, synthetic biology. Through the expansion of biological systems regulation and improving safety and efficiency of engineered systems, we can now develop novel and more efficient vaccines and antibodies, drug delivery methods, tissue and organ repair strategies as well as cell and gene therapy approaches.

2.A. Diagnostics

Access to easy-to-use, reliable, and time-saving methods for disease diagnosis has always been the ultimate goal of diagnostics and its associated fields. The arsenal of tools generated by gene synthesis is beginning to facilitate and expedite such attempts, and also bring many futuristic ideas into reality. Diagnostic tests take advantage of engineered RNA and DNA as well as their associated synthetic circuits. For example, engineered gene expression systems can now be coupled with synthetic RNA switches and linked to a colorimetric output detection, the combination of which results in a visual output only in the presence of a specific biomarker or pathogen. Such systems, printed on paper for portability, are currently being tested for detection of antibody resistance and viral infection[47, 48].

2.B. Gene Therapy

Gene therapy involves the delivery of therapeutic genes packaged within viruses into the diseased tissue of interest. Gene synthesis is helping advance gene therapy on several fronts: engineering efficient therapeutic genes, improving the specificity of gene delivery, and efficient expression of therapeutic genes in the patient. Cell or tissue specificity is achieved by expressing therapeutic genes from actively expressed cell or tissue-specific promoter elements whereas the combination of virus variant libraries and in vivo-directed evolution can identify the most efficient virus for gene delivery[49, 50]. This approach can also be combined with autologous cell transplantation for enhanced efficacy. The utilization of synthetic biology and its tools is allowing for the removal of stem cells from a patient, correction of defected genes with the use of programmable transcription factors, and then transfer of corrected genes back to the same patient with the help of engineered viruses. The CRISPR-Cas9 genome editing, in particular, is quickly growing into the main and most efficient choice for a variety of gene therapies for treating genetic diseases and cancer[51, 52]. The advantage of CRISPR is in its versatility, flexibility, and specificity: versatility in the use of an RNA molecule to specify a target rather than a protein domain, flexibility in designing libraries of transcription factors, and specificity in targeting the right sequence.

2.C. Cell Therapy

Cell therapy, which is now at the forefront of modern therapeutic interventions, is also benefiting from gene synthesis and synthetic biology. In particular, application of synthetic biology in Chimeric Antigen Receptor T-Cell (CAR-T) therapy is demonstrated to be successful in treating tumors that were not responsive to other forms of therapy. In this strategy, synthetic CAR libraries help identify the best epitope, T-cells are engineered to express the identified epitope *ex vivo*, and then engineered cells are transferred back into the patient to destroy the diseased target cells[53]. While the majority of existing CAR-based therapies target cancer cells based on the presence of a single cancer-specific antigen on any cell, emerging approaches are focused on improving specificity through the application of "AND logic gates". By requiring the presence of two antigens or one antigen and a small molecule drug, such systems ensure the targeting of specific cells only[14].

2.D. Vaccine & Therapeutic Antibody Development

Synthetic biology approaches that involve bioinformatics, gene synthesis, and mammalian or plant cell line production systems are giving rise to the production of the second generation of vaccines and therapeutic antibodies. Discovery and engineering of therapeutic antibodies is greatly advanced through the use of high-throughput screenings as well as DNA and mRNA display technologies powered by synthetic biology. Similarly, development of improved conventional as well as novel DNA- or RNA-based vaccines has been made possible through the application of synthetic biology tools[54, 55]. *In silico* design of novel immunogens, comprehensive viral sequence libraries, and RNA and codon usage modifications for improving antigen expression enable the development of vaccines with improved safety profiles, increased immunogenicity, and an enhanced breadth of immune responses[56, 57]. In particular, combinatorial application of several gene synthesis and synthetic biology approaches in the development of RNA-based vaccines offers promising therapeutics. These vaccines, which are used for conferring immunity against pathogens and cancers lack the ability to mutate the genome[58]. Therefore, they are preferred over DNA vaccines for safety concerns. The design of RNA-based vaccines through engineered immune cells is also foreseen to help overcome systemic immunosuppression and increase efficiency[59].

2.E. Regenerative Medicine

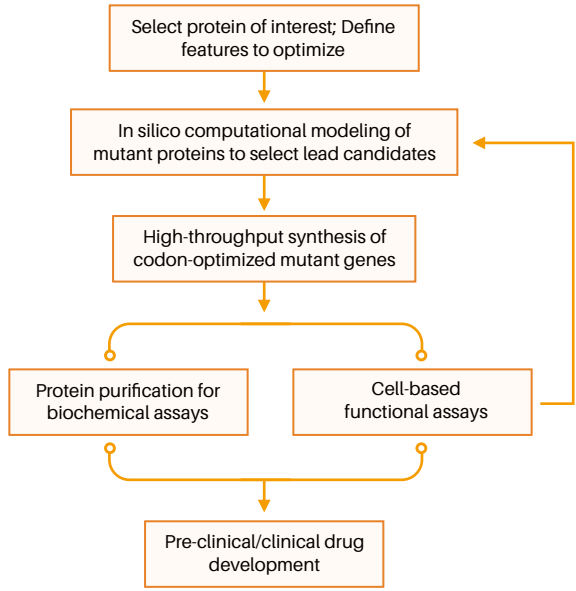
The emergence of the field of synthetic biology has coincided closely with the rise of regenerative medicine, also called tissue engineering. However, until recently the enormous potential of applying synthetic biology in tissue engineering was overlooked. Regenerative medicine's main goal - to regenerate damaged tissues through developing biological substitutes - has benefitted greatly from synthetic biology's tools[60]. This has been achieved through the biosynthesis and controlled release of therapeutic molecules, synthesis of scaffold materials, regulation of stem

cells, and programming cells that can organize themselves into novel tissues. DNA-based hydrogels as scaffold for tissue repair, synthetic multi-part signaling networks as patterning elements[61], and three-dimensional printing of living tissues capable of displaying cellular behavior and physiology[62] are a few examples of the application of synthetic biology in tissue engineering.

2.F. Drug Delivery

Tools generated by the gene synthesis technology are also being incorporated into the design of smart drug delivery systems to increase the drug targeting efficacy while reducing side effects and toxicities of payloads[63]. These systems, which take advantage of prokaryotic and eukaryotic cells, achieve a comparable therapeutic effect with much lower dosage of a drug, offer a less invasive route of administration with gastric enzyme-sensitive drugs, and provide a more cost-effective treatment through the *in vivo* production and delivery methods[64]. Another application is in developing prosthetic networks, which are synthetic circuits capable of sensing disease-relevant metabolites for coordinating diagnostic, preventive, and therapeutic responses[65, 66]. These networks are implanted in biocompatible genetically engineered encapsulated cells and delivered into the body. Upon sensing imbalanced levels of a metabolite or marker, such as high glucose levels, these cells can send a diagnostic signal, such as light, and release the right dose of a therapeutic drug, such as insulin.

Process of Engineering & Validating Novel Therapeutic Proteins



Expression of the E6 HPV oncoprotein, synthesized and cloned by GenScript, in 293 cells enabled Xu et al. to compile a comprehensive transcriptional and splicing signature of its expression towards better understanding of the cervical cancer pathogenesis.

A transcriptomic landscape of human papillomavirus 16 E6-regulated gene expression and splicing events. FEBS Lett, 2016.

3 Agriculture and Environment

Application of gene synthesis in plant sciences has far exceeded its original use in understanding plant cell biology, physiology and signaling pathways, or discovering plant-derived pharmaceuticals. Now, gene synthesis is enabling scientists to develop new and sustainable forms of energy, improve the quality and quantity of plant and animal-based sources of food, and remediate the detrimental effects of environmental pollutants.

3.A. Food & Sustainable Energy Production

Successful gene manipulation and transfer of recombinant or exogenous genes to laboratory animals in the early 1980s promised broader applications in plant and animal agriculture[67]. Decades later and with the help of gene synthesis, those promises are fulfilled not only at an exponential pace, but are also expanded into novel and creative directions. In plant agriculture, manipulation of existing genes and introduction of synthetic transgenes of plant, viral, or bacterial origin into a variety of crops have significantly improved crop yield through modulating photosynthesis and resistance to multitude of biological or environmental stress factors. Also, modulation of signaling pathways through gene synthesis has conferred novel nutritional values, and improved the natural health value and flavor of many plant-based sources of food[68, 69]. Transgenic plants are also extensively used for the production of antibiotics, growth hormones, enzymes, vaccines, and other therapeutic or diagnostic molecules[70]. Moreover, manipulation of pathways involved in the synthesis of starch, cellulose, hemicellulose, and oils has led to the generation of biofuels with minimum environmental footprint[71]. Similarly, application of gene synthesis in animal sciences allows for altering the genetics and physiology of livestock species to increase the yield and dietetic value of produced milk and meat[72].

GenScript codon-optimized and synthesized the scorpion toxin gene LqhIT2 to help Tianpei et al. confer pest resistance to rice.

Scorpion peptide LqhIT2 activates phenylpropanoid pathways via jasmonate to increase rice resistance to rice leafrollers. Plant Sci, 2014.

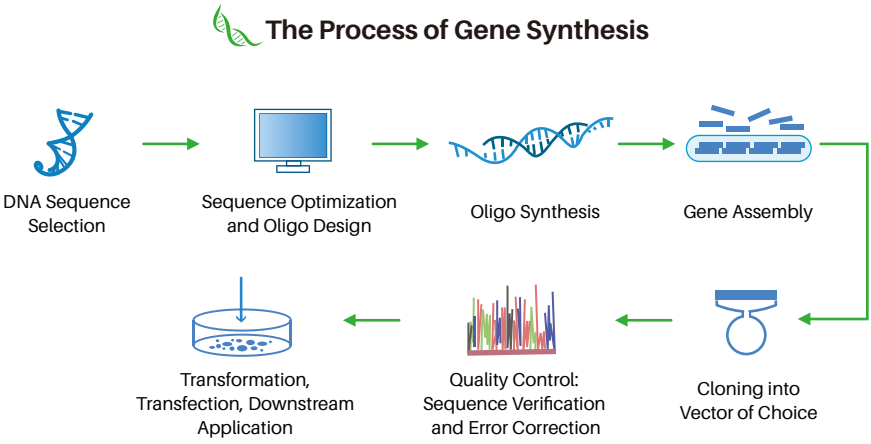
2.B. Bioremediation & Environmental Preservation

Conventional approaches for solving environmental pollution, caused by accidental spills or poor human management, are found to be ineffective and costly. Application of gene synthesis and its products, however, have facilitated the development of novel techniques that are more effective, faster, less expensive, and more environment-friendly. One method is bioremediation that is based on degradation or immobilization of organic and non-organic pollutants by engineered microorganisms. Using synthetic vectors, DNA and RNA libraries, and optimized or novel gene sequences, regulatory systems are built and engineered into microorganisms such that they could both sense and process contaminants. Functionality of these systems is based on the production and display of extracellular enzymes or proteins on the outer cell membrane that can monitor and catalyze or trap a specific category of contaminants[73]. A similar application of gene synthesis is also used to engineer organisms capable of producing chemical compounds that are biodegradable and less taxing on the environment in comparison to conventional sources of chemicals[74]. A prominent example is an engineered strain of E. coli with the ability to synthesize 1,4-butanediol as a precursor for the manufacturing of plastics, rubber, and solvents[75]. The production of this non-natural chemical, based on the construction of a novel pathway with no "blueprint" in nature using gene synthesis, helps preserve the environment through decreased dependency on oil-based chemicals. Development of engineered crops for the generation of biofuels or with modified physiological features with less dependency on water, carbon or nutrients along with bioengineering ecosystems are among the other benefits of applying gene synthesis in preserving the environment[74, 75].

Chapter Three

Technology

The chemical synthesis of genes and genomes involves four major steps. This process starts with identifying the application of the synthetic gene followed by the design and synthesis of short oligonucleotide sequences for further assembly to compile the whole gene. The third step is quality control to verify the sequence and the last step is to correct possible errors before the synthesized sequence is prepared for its downstream application. While advancements in chemistry and methodology have improved the efficiency of each step since the early years of the gene synthesis technology over 30 years ago, yet the core of each step has remained the same. This chapter provides a brief overview of the process behind the current, commercial gene synthesis technology.



1 Sequence Design & Optimization

The first step in synthesizing a gene of interest is designing the right sequence. In the design of a synthetic sequence two aspects need to be taken into consideration: (1) downstream application of the synthetic gene, and (2) the synthesis and assembly steps.

1.A. Downstream Application Optimization

Keeping the ultimate application of the synthetic gene in mind is an essential part of the design since placement of each nucleotide within the sequence can directly affect the success of the downstream application. The following is a list of examples that illustrates the importance of sequence design and optimization:

- In applications involving maximum heterologous protein expression levels, sequence design has to consider codon optimization. In the contrary, if the synthetic gene is to be used for studying endogenous gene regulation, codon manipulation should be avoided.
- When the synthetic sequence contains multiple segments, sequence design needs to ensure that throughout the entire coding region the reading frame is maintained.
- In cases where short flanking sequences are added to the gene sequence in order to facilitate later excision or recombination, careful attention needs to be given to avoid including or introducing restriction enzyme recognition sites or other sequences that might interfere with downstream workflow.
- Presence of biologically functional domains, such as cis-regulatory elements or RNase splice sites, may hinder *in vivo* assembly, maintenance, or expression of the synthetic gene. Therefore, sequences coding for such domains have to be avoided during the design step.

1.B. Synthesis Optimization

After finalizing the sequence design with regard to downstream applications, the sequence is analyzed to ensure a successful assembly. Given that synthetic genes are made up of assembled short oligonucleotide sequences of 40-200 bp, sequence analysis helps determine the best way to divide the whole gene into fragments that are easily synthesized and then seamlessly assembled (see step 3 below). In addition, some sequence features make the process of gene synthesis more challenging. These features include extremely high or low GC content, highly repetitive sequences, complex secondary structures, unstable structural elements, polyA stretches, and very long sequences over 1 kb.

Considering all the above-mentioned factors in the design of oligonucleotides can, therefore, be a confusing, time-consuming, and difficult process. Computational software for designing and optimizing oligonucleotides can significantly facilitate this process. Sophisticated algorithms incorporated in a variety of oligonucleotide design tools take all factors affecting downstream application, synthesis and assembly steps into consideration to ensure the synthesis of a functional sequence with minimum cost and turnaround time.

Oligonucleotide Design Tools

Design Tool	Advantages	Limitations
DNA Works http://helixweb.nih.gov/dnaworks/	Easy to use. Predicts the potential for oligo mishybridization and secondary structures.	Limited to PCR-based methods. Scores based on simulations may not correlate actual assembly success.
Gene2Oligo http://berry.engin.umich.edu/gene2oligo/	Simple user interface, designs oligos for both LCR and PCR-based assembly.	Limited to genes less than 1 kb in length.
GeneDesign http://baderlab.bme.jhu.edu/gd/	Breaks multi-kb sequences into ~500 bp blocks for initial assembly in separate pools.	Does not offer mishybridization analysis.
TMPrime http://prime.ibn.a-star.edu.sg	Designs oligos for both LCR- and PCR-based assembly. Produces the most homologous melting temperatures (Δ Tm < 3°C) and widest range of annealing temperatures (50–70°C).	Requires thorough understanding of many input parameters for correct submission; less user-friendly interface.

By considering a variety of critical factors involved in the different stages of protein expression, such as codon adaptability, mRNA structure, and various cis-elements in transcription and translation, GenScript's proprietary OptimumGene™ provides you with the ideal sequence to maximize protein expression in any host. Take the guess work out of optimization with GenScript's industry-leading gene optimization service.

2 Synthesis

In principle, DNA fabrication is oligonucleotide synthesis through a cyclical process in which nucleotide monomers are linked together in step-wise, chemical reactions. Synthesis is traditionally carried out in a solid phase in the form of a column, although the recently developed microfluidics and microarray platforms are increasingly becoming commonplace. The solid support, which is usually glass or polystyrene beads, serves as an anchor for the growing oligonucleotide chain. The first monomer is attached to the solid support through its 3′ OH (hydroxyl) group, then incoming monomers, based on the designed sequence, are added sequentially until the desired length is obtained and the chain is cleaved off of the solid phase.

The most commonly-used synthesis technique is the "phosphoramidite synthesis method", which was developed in the 1980s[77]and is now fully automated and robust. Whereas DNA synthesis in living cells always occurs in the 5′ to 3′ direction, the phosphoramidite synthesis proceeds in the 3′ to 5′ direction. Moreover, the building block nucleotide monomers used in this method are modified versions of the naturally occurring nucleotides, and are called "nucleoside phosphoramidites". Since natural nucleotides and their phosphodiester analogs are not reactive enough for a fast and high-yield rate of inter-nucleosidic linkage formation, using 3'-O-(*N,N*-diisopropyl phosphoramidite) derivatives improves the efficiency of synthesis. The functional groups of phosphoramidite monomers as well as any intermediate products generated throughout this cyclical method, are conjugated with a variety of protecting groups to avoid undesired side reactions. For example, the 5'-hydroxyl group of the first and all incoming monomers is protected by an acid-labile 4,4'-dimethoxytrityl (DMT) group.

The phosphoramidite synthesis method involves four main steps, which are cyclically repeated until the desired oligonucleotide chain is synthesized. These steps are:

1. **Deprotection:** The protecting DMT group on the first monomer attached to the solid phase is removed by a mild acid wash to expose the 5′ hydroxyl group for the coupling reaction;

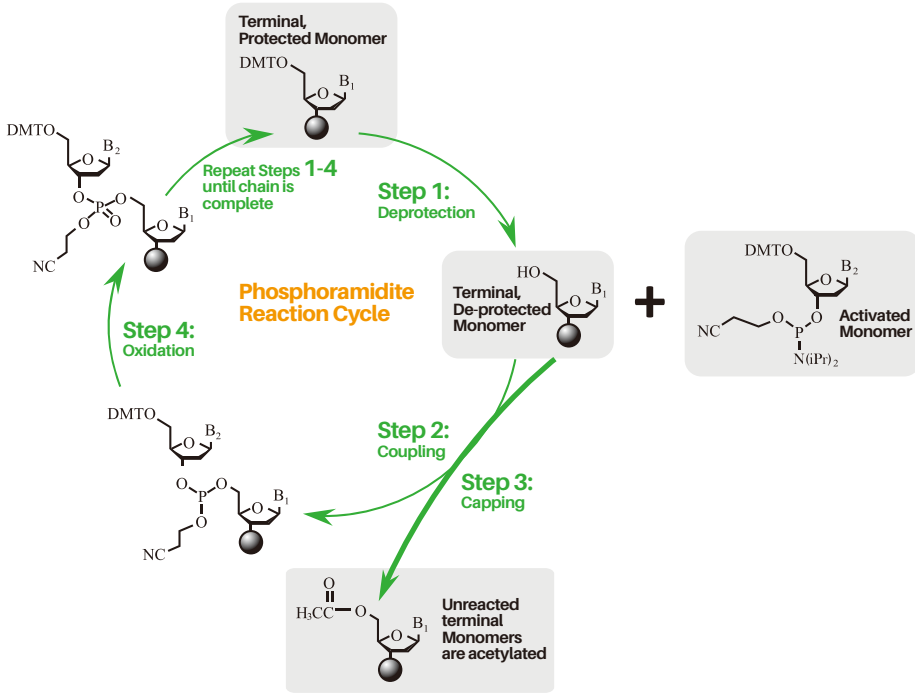
2. **Coupling:** The second monomer with its DMT-protected 5′ hydroxyl group and 3′ hydroxyl activated by the conjugated phosphoramidite group enters the reaction. Coupling occurs when the 3′ hydroxyl group of the second monomer forms a phosphate triester bond with the 5′ hydroxyl group of the first nucleotide;

3. **Capping:** Uncoupled 5′ hydroxyl groups of any unreacted nucleotides are blocked by acylation to prevent later growth of an incorrect sequence. Typically, acetic anhydride and dimethylaminopiridine are used which will have no effect on terminal nucleotides that are protected by DMT;

4. Oxidation: By adding iodine, the relatively unstable phosphite triester inter-nucleotide bonds are converted into the phosphodiester bond that forms the familiar backbone of DNA.

Once the synthesis of the desired chain is complete, the protecting groups are removed by treatment with a strong base such as ammonium hydroxide, and the 5′ end of the oligonucleotide is phosphorylated. Prematurely terminated strands are also removed by purifying the eluted product via gel electrophoresis and cutting out the band with the correct length. Synthetic oligonucleotides are usually further characterized by sequencing or are simply subjected to mass spectrometry to verify the predicted molecular mass.

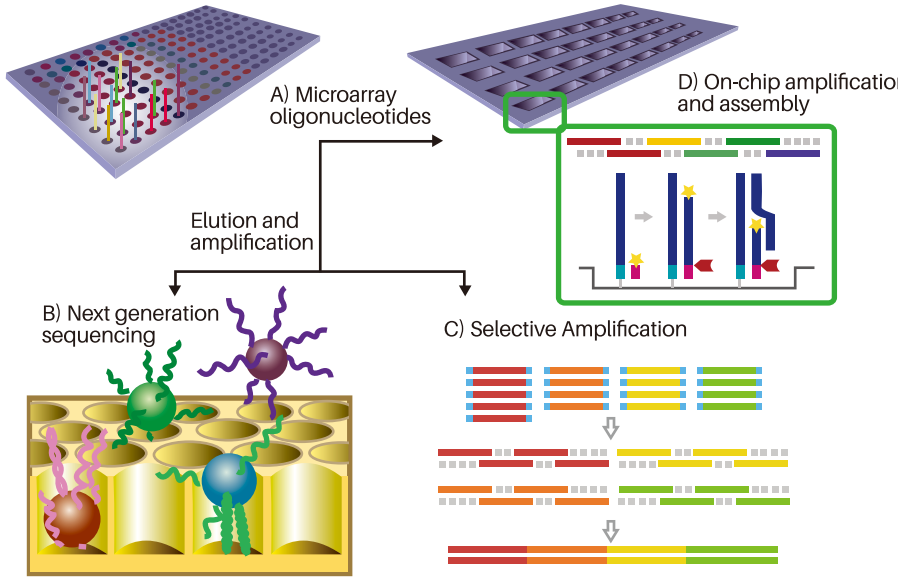
Phosphoramidite Oligonucleotide Synthesis Reaction Cycle



Solid phase phosphoramidite chemistry can generate oligonucleotides with lengths < 100 nucleotides with coupling efficiencies of almost 99%[78]. The automated process allows the synthesis of 96-384 oligonucleotides simultaneously at 10-100 nmol scales[79]. Over the last few decades, novel modifications in each step of the synthesis process as well as improvements in raw materials, have led to the development of improved methods to decrease the amount of starting materials, reagents, and waste needed to produce oligonucleotides, resulting in a more affordable and productive

process with both faster speeds and greater yields. Emerging microfluidics and microarray platforms used in high-throughput gene synthesis are among such advancements[80]. Microarray technologies are based on spatially confining reagents and directing reactions with very fine spatial and temporal control. Such methods rely on using inkjet printers to dispense picoliter reagents to specific locations on a silica chip, controlling the deprotection step with light-activated photochemistry in a microfluidic system. These technologies also use programmable microelectrode arrays to direct redox reactions at desired spots. In practice, however, these methods can create error through imprecision and can introduce new costs, such as the expense of manufacturing unique photomasks required for photolithography platforms.

High-Throughput Array-Based Gene Synthesis Technology



GenPlus™ High Throughput Gene Synthesis Technology is GenScript's latest breakthrough: a proprietary technology platform combines parallel synthesis and automation to enable an unprecedented capacity of 100 Mbp per month; the equivalent of up to 100,000 genes, or 20 bacterial genomes! Tackle gene synthesis projects of any size with GenScript's industry-leading gene synthesis expertise today!

3 Assembly

The chemical synthesis of oligonucleotides can efficiently generate sequences of up to 120-150 nucleotides. However, since the synthesis of oligomers with longer lengths are associated with lower yield and higher error rates, it is recommended to fabricate long sequences through assembling shorter sequences. In the past several decades, a number of assembly methods have been developed. These techniques are ligation-mediated, PCR-mediated, or a combination of both methods. Increasingly, recombination-mediated methods are also used to simultaneously assemble and clone long DNA sequences for increased efficiency and reduced error generation.

Main Assembly Methods for Synthetic Genes & Genomes

Oligo Assembly Methods	Reference
POLYMERASE-BASED	
Dual-Asymmetric (DA) PCR	Sandhu *et al*. (1992)[88]
Overlap Extension (OE)	Horton *et al*. (1993)[89]
Polymerase Cycling Assembly	Stemmer *et al*. (1995)[90]
Asymmetric PCR	Wooddell & Burgess (1996)[91]
Thermodynamically-Balanced Inside-Out (TBIO)	Gao *et al*. (2003)[81]
Two-Step (DA+OE)	Young & Dong (2004)[92]
Microchip-Based Multiplex Gene Synthesis	Tian *et al*. (2004)[93]
One-Step Simplified Gene Synthesis	Wu *et al*. (2006)[94]
Parallel Microfluidics-Based Synthesis	Kong *et al*. (2007)[95]
Single Molecule PCR	Yehezkel *et al*. (2008)[96]
TopDown Real-Time Gene Synthesis	Huang *et al*. (2012)[97]
LIGASE-BASED	
Shotgun Ligation	Eren & Swenson (1989)[98]
Two-Step Ligation and PCR	Mehta *et al*. (1997)[99]
Ligase Chain Reaction	Au *et al*. (1998)[100]
Brick-Based	Kelly *et al*. (2009)[101]
RECOMBINATION-BASED	
Sequence-and Ligation-Independent Cloning (SLIC)	Li and Elledge (2007)[102]
Transformation-Associated Recombination	Gibson *et al*. (2008)[103]
BioBrick Assembly in *E. coli*	Ho-shing *et al*. (2012)[104]

The following is a list of major categories of early and current assembly methods with a few example techniques in each group.

3.A. Polymerase-Mediated Assembly

This category of assembly techniques are the most commonly used, and rely on the polymerase chain reaction to extend and assemble overlapping oligonucleotides into double-stranded fragments by cycling in a non-exponential process[79]. Assembly in these techniques can be single-step in which the desired gene product is assembled in either a single reaction or in multiple enzymatic reactions. Careful application of oligonucleotide design software packages is essential in the success of the assembly step since each method is based on unique primer sets. Polymerase Cycling (or Chain) Assembly (PCA)-based method is one standard technique in this category. In this templateless PCR, all of the single-stranded oligonucleotides are combined into a single tube and are subjected to thermocycling with the use of the outermost primers to amplify the full-length sequence. The success of this method depends upon the accurate synthesis of oligonucleotides designed to possess sufficient regions of overlap, similar melting and annealing temperatures, and minimal opportunities for mis-hybridization. In another multi-step method, called Thermodynamically Balanced, Inside-Out (TBIO) assembly, overlaps between primers are balanced in order to obtain uniform annealing temperatures across the assembly. Besides this novel approach to primer design, primers are added in a concentration gradient in which the innermost pair has the lowest concentration and the outermost pair the opposite. The selective pressure for amplification will lead to the assembly of longer products, favoring the fabrication of the full-length, desired gene[81].

3.B. Ligation-Mediated Assembly

Ligation-based assembly techniques take advantage of the enzymatic activity of bacterial ligase for joining complementary overlapping strands to produce larger DNA fragments. This category represents the earliest methods used in gene synthesis, including the genes for yeast alanine tRNA and human insulin A synthesized in the late 60s and 70s, respectively[82]. These methods were initially carried out sequentially, but access to higher quality oligonucleotides, application of thermostable ligases such as *Pfu* DNA ligase, and development of methods for producing and selecting for circular DNA have enabled a one-pot assembly of gene fragments[79]. An example method in this category is the Ligase Chain Reaction (LCR) assembly, which involves a temperature cycling reaction similar to PCR. In this technique, oligonucleotides are synthesized in such a way that their 5′ end is phosphorylated and each contain overlapping sequences that span both strands of a desired DNA duplex. Once these oligonucleotides are mixed together, they are heated to denature, and then cooled to promote both the annealing of oligonucleotides as well as their ligation by the *Pfu* ligase. The newly assembled intermediate products can then serve as templates for additional ligation

events until the desired gene has been assembled. In case of longer DNA fragments, a final PCR step is often included to ensure the amplification or stitching of the shorter fragments. Another example in this category takes advantage of a combination of several enzymes, such as T5 exonuclease, Taq ligase, and Phusion DNA polymerase to assemble DNA fragments up to several hundred kilobases *in vitro*[83]. In this approach, T5 exonuclease first digests the 5′ end of overlapping DNA duplexes to generate single-stranded, complementary fragments. These fragments then serve as primers for Phusion DNA polymerase which fills in the empty overlapping areas. Lastly, nicks in the assembled sequence are sealed with a thermostable Taq ligase.

Application of a thermostable enzymes in ligation-mediated assembly techniques allows for high-temperature annealing and ligation steps that improve the hybridization stringency and the likelihood of obtaining correct final sequences, which may offer an advantage over PCA. However, this method becomes in efficient as the number of oligonucleotides increases; limiting its usefulness to genes shorter than 2 kb[84]. Furthermore, the use of standard-purity oligonucleotides can introduce point deletions resulting from incorporation of incorrectly synthesized oligonucleotides. This can be avoided by using gel-purified oligonucleotides, high-quality synthetic oligonucleotides, or by performing side-directed mutagenesis to correct errors after assembly, cloning, and sequencing.

3.C. Recombination-Mediated Assembly

This category of assembly methods, performed both *in vitro* and *in vivo*, is most suitable for the fabrication of very large DNA molecules and combines the assembly step with downstream cloning. The principle is founded on assembling small DNA fragments containing specific sequences with homology to other fragments within the designed gene and a plasmid vector. For example, in Sequence- & Ligation-Independent Cloning (SLIC), which is a method of *in vitro* homologous recombination and single-strand annealing in one reaction, a T4 DNA polymerase is employed to assemble up to 10 gene fragments via simultaneous incorporation into a plasmid vector[85]. Using PCR extension to introduce flanking regions of sequence homology, synthetic oligonucleotides are prepared for SLIC; this facilitates recombination of fragments without any sequence restrictions or the introduction of restriction enzyme sites that will produce permanent seams. The exonuclease activity of T4 DNA polymerase generates single-stranded DNA overhangs in the insert and vector sequences. Homology regions are annealed *in vitro* and undergo gap repair after transformation into *E. coli.*

Assembly of mega base-length DNA, however, requires *in vivo* recombination assembly, which is usually carried out in *Bacillus subtilis, E. coli*, or yeast. Using a domino method, for example, 4–6-kb DNA fragments termed “dominos”, are synthesized such that the ends of the DNA contains regions of homology on one end to the *B. subtilis* genome and on the other end to another domino fragment. Once these pieces are constructed in conventional *E. coli* plasmids carrying antibiotic resistance genes, the

entire designed DNA sequence can then be assembled in the *B. subtilis* genome using multiple rounds of transformation and selection[86]. Using high efficiency recombination enzymes, such as lambda Red in bacteria as well as employing the intrinsic high efficiency system of homologous recombination in yeast and its ability to replicate large DNA molecules, a variety of recombination-based methods have been developed.

In vivo homologous recombination in yeast facilitates the assembly of very long synthetic DNA sequences either from "bricks" of 500-1,000 bp produced through *in vitro* methods, or directly from short oligonucleotides. Recently, the creative work of Gibson and colleagues has demonstrated that the yeast *Saccharomyces cerevisiae* can take up and assemble at least 38 overlapping single-stranded oligonucleotides and a linear double-stranded vector in one transformation event. These oligonucleotides can overlap by as few as 20 bp and can be as long as 200 nucleotides in length to produce megabase-sized synthetic DNA molecules[32, 83]. This method employs three enzymatic activities in a single, isothermal reaction: (1) 5´ exonuclease activity chews back the 5´ end of DNA fragments and exposes the complementary sequence for annealing, (2) the 3´ extension activity of a DNA polymerase fills in the gaps on the annealed regions, and finally (3) the DNA ligase activity seals the nick and covalently links the DNA fragments together for final placement of the synthetic gene(s) in the destination vector. Integration of multiple methods with Gibson Assembly is now enabling the modular construction of large, multigene plasmids for mammalian cell experiments, which is widely used in synthetic biology[87].

Main Assembly Design Platforms

Platform	Features
J5	• Designs flanking homology sequences or overhangs for each assembly piece • Enables scar-less multipart and combinatorial DNA construction • Applicable to broad classes of DNA construction methodologies • Compatible with many assembly protocols • Optimizes for cost and parts re-use • Requires carefully-edited raw data • Graphical user interface • Licensing restrictions
Raven	• Designs oligonucleotide sequences to prepare parts for assembly • Generates optimized assembly plans • Compatible with most popular assembly protocols • Optimizes for cost and parts re-use • Requires carefully-edited raw data • Graphical user interface • Licensing restrictions
Pydna	• Designs DNA constructs and assembly plans • Useful for complex or combinatorial DNA molecule construction • Compatible with Gibson and recombination-based assembly methods • Free and open source • Flexible and extensible • Bottom-up approach allows gradual building and testing for complex projects • Requires basic knowledge of Python coding language
Gibthon	• Designs primers for Gibson Assembly • Web-based and interactive • Free and open source
BioPartsBuilder	• Designs DNA constructs and assembly plans • Compatible with combinatorial assembly methods • Web-based and interactive • Free and open source

4 Sequence Verification & Error Correction

Once the desired synthetic gene has been assembled, it needs to be verified to ensure the final sequence matches the one originally designed. Due to inherent potential for error in each step of the gene synthesis process, the assembled DNA product contains a mixture of perfect and imperfect sequences. Internal insertions, deletions, and truncated fragments are the result of inefficient coupling, unwanted DMT cleavage, incomplete capping, or deprotection during the cyclical synthesis process. Improper annealing during oligo assembly can also introduce heterogeneity in the final pool of synthetic gene products. Also sequences harboring mutations must be identified and removed from the pool or corrected. To screen synthetic products and correct potential errors, several methods have been developed:

- **Oligonucleotide and/or gene fragment purification:** Since the quality and purity of the oligonucleotides used in the process of chemical gene synthesis have the potential for errors, many gene synthesis protocols rely on purification of full-length oligonucleotides using PAGE or agarose gel electrophoresis prior to assembly. A similar approach is also used for selecting correct gene fragments before assembly. This approach, though helpful, is time- and labor-intensive and does not identify or correct for substitutions, small insertions, and deletions.
- **Cloning and sequencing:** Typically, newly synthesized sequences are cloned into a plasmid vector and subjected to Sanger sequencing. Sequencing primers that bind to vector regions flanking the gene insert ensure correct sequencing of the ends of the synthesized gene insert. Further, plasmid DNA can be clonally amplified to create a homogeneous pool of DNA with the correct sequence. This approach is specifically useful for larger DNA molecules in which segments of the gene or genome, called synthons, are first cloned and sequenced before the whole gene/genome is assembled[86]. This approach can be expensive, time consuming, and difficult to automate.
- **Reading frame selection:** For synthetic protein-coding sequences, functional selection can be used to prescreen and enrich mutation-deficient sequences. Since many of the accumulated errors are single point mutations leading to frame shift and premature translation termination, the synthetic genes can be cloned into a frame-shift selection vector that will express a fusion reporter protein as a selectable marker, such as antibiotic resistance or fluorescence markers. Selection is, hence, based on the survival of intact genes transformed into bacteria and plated on selective media.
- **Mismatch binding and cleavage:** This approach takes advantage of mismatch repair endonucleases, which bind to error-containing sequences and affect their electrophoretic motility. Error-free DNA fragments migrate faster on an electrophoresis gel, and can be eluted and further assembled. The efficiency of this approach is, however, dependent on the capability of the enzyme. For example, the widely-used *E. coli* MutHLS is not very effective for substitutions other than G-T, A-C, G-G, A-A.

- **Site-directed mutagenesis:** Following gene assembly or sequencing, primers containing the targeted correction can be used to directly correct errors through site-specific mutation. These primers can be the original ones designed for the synthesis and assembly of the gene(s) or newly synthesized. This approach, although dependent on one or two extra assembly steps, is very efficient in correcting errors.
- **NGS-based error correction:** Advances in Next Generation Sequencing (NGS) have recently been employed to screen and then select for perfect sequences at both the oligonucleotide and gene level. Using array sequencing as is, or integrated with tags and biomarkers, all variations of this approach help expedite screening and selection of correctly-synthesized DNA molecules. While these techniques still need improvements in sequencing and length-accommodation, they are promising in that they enable simultaneous correction without the need to separate each gene assembly into individual reactions[79].

Since all error detection and correction strategies are time-consuming and costly, especially for long or complex sequences, efforts continue to improve the accuracy of initial oligonucleotide synthesis and assembly steps. Recent strategies that rely on synthesizing and assembling fragments in a single reaction is one example in which higher percentage of correct and ready-to-use synthetic products of very long sequences are obtained.

5 Downstream Application Preparation

The correct final product of the chemical synthesis process can be used directly in a few downstream applications. However, for the majority of applications, synthetic genes and genomes need to be cloned and propagated. Development of recent synthesis techniques in which assembly and cloning are combined can expedite the preparation of the product for downstream applications, but sub-cloning and propagation may still be required for alternative applications.

5.A. Cloning

Depending on the specific downstream application, synthetic genes can be cloned into plasmid vectors for transfection or electroporation into cells or viral vectors (such as adenoviral, retroviral, or lentiviral vectors) for transduction into cells or live animals. Cloning methods can also vary from traditional, multi-step techniques to more recent, one-step methods. Currently, main cloning methods include restriction enzyme cloning, TOPO® cloning, Gateway® recombination cloning, isothermal assembly reaction or Gibson Assembly®, and type II S assembly, such as Golden Gate. The choice of cloning strategy depends on the synthetic gene and its intended application goal. Alternatively, specific features and requirements of each cloning method dictate specific required modifications in the synthetic genes before cloning can occur.

Cloning Vectors

Type	Insert Size (kb)	Host	Delivery Method	Features
Bacterial plasmid	0.1-15	*E. coli*	Transformation	Circular; Double stranded; Small size; Single restriction site; Contains suitable markers for selection; Multiple cloning sites
Bacterio-phage	10-300	*E. coli*	Transduction Transduction or	Linear; Single stranded; More efficient than plasmids
Cosmids	35-45	*E. coli*	transformation	Behaves as both plasmid and phage vectors; High transformation efficiency
BAC	100-300	*E. coli*	Electroporation	Linear or circular; Suitable for large eukaryotic genomes and genome analysis
YAC	200-2000	*S. cerevisiae*	Transformation	Linear or circular; Low transformation efficiency; Suitable for very large eukaryotic genomes and mapping complex eukaryotic chromosomes
Human Artificial Chromosome	>2000	*Mammalian & human cells*	Microcell-mediated chromosome transfer	Linear or circular; Suitable for very large eukaryotic genomes, gene delivery and insertional mutagenesis

Main Cloning Strategies

Strategy	Methodology	Efficiency	Features
Restriction Digest Cloning	Restriction enzyme digestion of insert and vector	Variable	Classic, cheap and affordable
			Convenient for subcloning
			Generates blunt or sticky ends
			Can clone only one fragment per reaction
			Compatible with many vectory types
			Easy to design
			Leaves a scar or seam
			Can be directional or not
			Sequence limitations
T/A Cloning	Hybridization of single base overhangs in insert and vector	Variable	Suitable for sequencing
			Can clone only one fragment per reaction
			Amenable to high-throughput
			Easy to design
			Leaves a single base insertion scar
			Not directional
			Multi-fragment cloning not straightforward
			Not compatible with many vectors
			Time-consuming preparation and costly
			Requires an entry vector
Gateway Cloning	Recombining "att" sites in insert and donor vector	~95%	Easy and time-saving strategy for subcloning
			Can clone several fragments per reaction
			Leaves an "att" site scar
			Can be directional or not
			Fast and accurate
			Compatible with many vectory types
			Relatively easy to design
			Requires an entry vector
			Time-consuming initial set up

Strategy	Methodology	Efficiency	Features
Gibson Assembly & Cloning	Isothermal assembly of homologus overlapping sequences	~90%	Can clone several fragments per reaction with high level of control
			Leaves no scars or seams
			Directional
			Compatible with many vectory types
			No sequence limitations
			One-step assembly and cloning
			Easy for multiple DNA manipulations, such as mutagenesis
			Fast
			Difficult to design
			High cost of synthesizing overlapping primers
Ligation Independent Assembly & Cloning	Complementary overhangs between insert and vector	~95%	Compatible with many vectory types
			Low cost
			Fast
			Limitations on sequence modifications
Golden Gate Assembly & Cloning	Restriction enzyme (Type II S-dependent) digestion	~95%	Can clone several fragments per reaction
			Leaves no scars or seams
			Directional
			Re-ligation prone
			Suitable for generating combinatorial library
			Less sequence-independent
			Difficult to design

Integration of advanced gene synthesis and cloning strategies are now providing scientists with "off-the-shelf" products to facilitate and expedite research. These products contain the desired gene sequence(s) already cloned in a variety of vectors for direct ordering and application in an experiment. Products containing the Open Reading Frame (ORF) of a gene in wild or mutant types, offered as ORF Clones, which are designed and ordered through virtual molecular cloning platforms, are such ready-to-transfect or -express products. Pre-fabricated gene fragments corresponding to functional or regulatory elements of a gene and its ORF can also be ordered for direct use in an application or cloning for future applications.

5.B. Propagation

Synthetic genes and genomes can be propagated through the amplification of the plasmid they are cloned in. This procedure, however, may be challenging for several reasons: some "low copy number" plasmids simply don't amplify well in commonly-used bacterial propagation hosts; long genes are often diffcult to maintain and propagate because of the energetic burden on the host cell; some genes may alter the physiology of their hosts, creating aberrant culture temperature requirements or other conditions that may be difficult to identify and accommodate; and genes may be toxic or unstable in certain host cells. One solution is to screen a number of cell lines to find the best host for propagation. While this strategy can be time consuming, but it may be necessary for certain ii sequences.

GenScript's **Express Plasmid Prep Services**, in both research and transfection grade packages, can accommodate any size and quantity of DNA preparations to help you achieve highly efficient cell transfection for DNA vaccine development, antibody production, and other preclinical research projects.

Chapter Four

Perspectives

Biosafety

Throughout the history of genetics research, numerous regulatory boards have addressed the safety concerns of researchers, citizens, and political leaders regarding genetic technologies. The potential for genetic technologies to be used for harm has always been a major point of discussion with each new scientific advancement. In 1974, the National Institutes of Health formed Recombinant DNA Advisory Committee (RAC) to review recombinant DNA safety protocols and address public misgivings about newly acquired scientific and technological powers to manipulate genetic materials. Although gene synthesis allows revolutionary flexibility for researchers, it poses no qualitatively different risks than those that have already been regulated for almost half a century since the advent of recombinant DNA technology. In 2009 the International Gene Synthesis Consortium (IGSC) was established to work closely with the research community and the Presidential Commission on the Study of Bioethical Issues to evaluate and manage biosafety risks. The IGSC has developed industry-wide protocols to screen all synthetic gene orders against databases maintained by national and private agencies to identify regulated pathogen sequences and other potentially dangerous sequences in order to prevent their misuse.

GenScript is proud to be a founding member of the **International Gene Synthesis Consortium** (**IGSC**), whose member companies represent more than 90% of the worldwide gene synthesis capacity.

Future

Gene synthesis technology has revolutionized both our understanding of how DNA functions as the blueprint of life and our ability to manipulate DNA for experimental, medical, and industrial purposes. While the capabilities and speed of gene synthesis have steadily increased, its cost has dropped from $10 per bp to $0.09 per bp and below over the last decade; in parallel with advances in DNA sequencing and chip-based bioassays. Further improvements in automation, assembly, error correction, high-throughput platforms, and cost-effectiveness will further allow gene synthesis to

be not only a valuable, but also an indispensable tool for a widening range of scientific disciplines and economically important applications. The energy and agriculture industries will increasingly rely upon gene synthesis to solve problems related to environmental protection and food supplies. DNA-based vaccines and bioengineered antibodies will become more widely used in medicine. Designer organisms housing customized metabolic pathways will become commonplace. Tinkering with the cellular circuitry to modulate biosynthetics, including synthetic sensors and regulators, will continue to be a focus of synthetic biology research. The potential environmental and social impacts of integrating synthetic genes into endogenous systems will kindle ongoing discussions among the research community, biotech industry, governmental and international regulatory bodies. Without a doubt, the coming decade will present exciting opportunities for gene synthesis to continue to fuel innovation in life sciences.

Chapter Five

Resources

Tools

To help facilitate the design and application of your synthetic genes, GenScript has developed a comprehensive set of powerful tools. The following introduction familiarizes you with the scope of each tool before you are ready to use them online through our secure and encrypted system at *https://www.genscript.com/gene_technical_resources.html.*

+ GenSmart Design

Helps you integrate all you need to design a DNA construct into 1 of the 3 simple, drag-and-drop operations: insertion, fusion and replacement. This smart construct design interface requires no installation and can automatically annotate your DNA sequence and draw plasmid map.

+ GenSmart Codon Optimization

Enables you to optimize the design of wild type or recombinant gene sequences towards higher expression in prokaryotic and mammalian expression systems. Developed based on population genetics and immunology theories, this free tool takes more than 200 factors involved in gene expression into consideration to maximize the chance of obtaining a functional and active protein.

+ Rare Codon Analysis Tool

Determines the codon usage frequency and distribution of your target sequence. It can help you decide if your sequence needs to be optimized for heterologous gene expression.

+ Restriction Site Map Analysis Tool

Analyzes your target sequence for restriction sites and displays where restriction enzymes will cut.

+ Gene Mutagenesis Designer

Facilitates the design of point mutations in your target sequence. Simply submit the full sequence, desired expression system, and location of each mutation to obtain the right mutated sequence.

+ Oligo Calculation Tool

Calculates melting temperature (Tm), molecular weight (MW), and millimolar extinction coefficient (OD/μmol) and μg/OD of oligonucleotides.

+ siRNA Target Finder

Identifies proper sequences within your sequence that would be efficiently targeted by synthetic sequences for gene silencing in your target organism or organ.

+ siRNA Construct Builder	Helps you build small hairpin inserts for use as siRNA expression vectors.
+ ORF Clone Vector List	Helps you find the right expression vector for the ORF clone of your choice by considering factors such as the expression host, promoter, bacterial selection, copy number, or epitope tag.
+ Sequence Manipulation Suite	Provides a collection of JavaScript programs for generating, formatting, and analyzing short DNA and protein sequences.

Design Considerations

To help you design functional synthetic genes and solve problems related to gene synthesis and application, consider the following guide in conjunction with our Frequently Asked Questions page at our Gene Technical Resource Center: *https://www.genscript.com/gene_technical_resources.html.* Our Ph.D.-level technical account managers are also available to answer your questions 24/7.

General Considerations

Advanced and thorough preparation is key in obtaining a functional synthetic nucleic acid product. Answering the following questions helps to prepare for your project:

- What is your specific scientific question?
- How does a synthetic gene help answer your question?
- Are you looking at the function of a gene or how it is regulated?
- Do you need the complete gene sequence or part(s) of it?
- Is it necessary to include regulatory elements of your gene of interest in the synthetic construct?
- Do you need the wild-type or mutated form of the gene?
- Do you need to have more than one gene sequence in your construct?
- Do you need to have the expressed, purified product of the synthetic gene(s) or assessing its cellular function suffice?
- Do you need to include reporter tags or molecules in your construct, and if so, where does it need to be placed to avoid interference with gene function?
- What biological system, prokaryotic or eukaryotic, can provide you with the most functional product?

- What synthesis and assembly methods are the most appropriate to employ for obtaining a functional synthetic product?
- What are the requirements for your choice of synthesis and assembly methods to ensure proper oligonucleotide design for a successful process?

Oligonucleotide Design

- Designing oligonucleotides for the synthesis of long DNA sequences or genomes can be a time-consuming, difficult, and confusing process. Make sure to know the details of your target sequence(s) and follow the specific requirements of your selected assembly method and cloning strategy.
- Particular aspects of the desired sequence(s) can interfere with gene synthesis and assembly. Pay attention to codon usage, GC content, restriction enzyme sites, and potential for secondary structures within your desired sequence(s). Most of these properties can be eliminated by juggling codon usage in a sequence.
- DNA repeats in the target sequence may occur in many forms including direct, inverted, palindromic, or tandem repeats. Presence of repetitive DNA sequences can lead to large deletions in homology-directed assembly reactions. In such cases, it is best to use restriction enzyme-based assembly strategies to avoid sequence-specific annealing reactions. Repeated sequences may also pose challenges for correct annealing or ligation of oligonucleotides during assembly.
- GC content determines the stability of DNA strands and thus the melting temperature. All in vitro assembly methods, both ligase and polymerase based, rely upon the melting and annealing of oligonucleotides and as a result require that oligonucleotides have equal melting temperatures.
- Improper hybridization or intramolecular binding, such as hairpin structure formation, can be avoided through careful design of oligonucleotides, including the overhangs, or primers that were used for assembly.
- To avoid introducing special flanking sequences in recombination-based assembly methods, introduce 15 bp with sequence homology to the linearized vector through PCR extension of the gene insert. This will facilitate homologous recombination without adding unwanted bases.
- Since cellular intermediates are a component of gene fabrication strategies, presence of toxic genes can interfere with gene fabrication. Make sure to check their presence in your target sequence(s). If they exist, avoid them by omitting their start codon or introduce a frameshift mutation to disrupt their activity. Once the synthetic gene(s) are fabricated, you can repair the mutation through site-directed mutagenesis.
- While synthesis platforms utilizing phosphoramidite chemistry have very low error rates, errors do accumulate as strand length increases; therefore, it is recommended to use oligonucleotides with lengths of 40-200 bp. The optimal oligo length depends upon the assembly method that will be used, the complexity of the sequence, and your preferences.

- As a general rule, shorter oligonucleotides may have a lower error rate, but will be more expensive to synthesize because more overlaps will be required. Longer overlaps increase the likelihood of correct assembly by decreasing the rate of nonspecific annealing.
- Oligonucleotides should be adjusted for even lengths and equal melting temperatures. Pay attention to parameters of oligonucleotide design tools if you choose to use them. Each tool has its own default parameters; for example, Gene Design uses default settings of 60 bp oligonucleotides with 20 bp overlaps; values which typically work well with yeast and mammalian sequences which have about 40% GC content.

GenScript's **OptimumGene™** is the industry-leading gene optimization service that takes multiple variables influencing gene expression into account in order to provide you with the ideal sequence for your research needs.

Plasmid Prep & Cloning

- Depending on your research needs, the quality of plasmid obtained from your gene synthesis project may need varying levels of quality control. Key factors are A260/280, residual RNA, endotoxin levels, and supercoiling percentages.
- It is essential for successful cloning that the plasmid purity is high. DNA absorbance as measured by a spectrophotometer can be used to determine the purity after plasmid purification: DNA, proteins, and solvents absorb at 260 nm, 280 nm, and 230 nm, respectively. The ratio of absorbance at 260 and 280 nm is used to assess the purity of nucleic acids. An $OD_{260/280}$ ratio of >1.8 and an $OD_{260/230}$ ratio of 2-2.2 are indicative of pure for DNA samples.
- Residual RNA in a plasmid prep can interfere with downstream applications. It is recommended to remove RNA with DNase-free RNase or use a Sephacryl S-300 column.
- Supercoiled plasmids have a better transformation efficiency than plasmids that are relaxed. Relaxed plasmids are transformed with 25% less effciency than supercoiled plasmids. Thus, higher levels of supercoiling in a plasmid prep results in a higher subsequent transformation effciency.
- Endotoxins are molecules that are comprised of a lipid and a polysaccharide. They are can be found within a bacterial cell and are released when the bacterial cell wall is destroyed (during the plasmid prep procedure). The removal and protection against contamination by these lipids are key in certain applications to avoid contaminating experiments and to avoid toxicity. Here are some applications that require lower endotoxin levels:

- Transfection and co-transfection of mammalian cells
- Antibody production
- Gene vaccine and gene therapy studies
- Protein production
- Animal studies

Bundling **Gene Synthesis Services** from GenScript with **Express Cloning Services** offers you free access to over 150 popular vectors for cloning your perfectly-synthesized sequence into mammalian, bacterial, yeast, or insect expression systems.

Storage & Handling

- After receiving the vials containing plasmids, make sure to store them at 4°C for short-term and frequent use, or at -20°C for long-term storage.
- Keep vials sealed until ready to use.
- Prior to use, centrifuge the vial at 6,000 x g for 1 minute at 4°C. Open the vial and add 20 µl of sterilized water to dissolve the DNA. Close the lid and vortex the vial for 1 minute. If necessary, heat the solution at 50°C for 15 minutes to dissolve the DNA.
- Transformation of the plasmid DNA can be performed directly after completion of the steps above. Verify the plasmid sequence after each subcloning and transformation step, unless you obtained your plasmid from a trusted vendor who ensures sequence verification.
- If you received your plasmid in glycerol, avoid repeated thawing and re-freezing of the glycerol stocks since it may reduce bacterial viability.

By taking advantage of GenScript's **VectorArk Storage and VectorArk Services**, you can securely store your vectors and customized plasmids for future cloning projects for free. Leave the worry of losing your precious clone behind!

References

1. Michelson, AM. and Todd, AR. *Nucleotides part XXXII. Synthesis of a dithymidine dinucleotide containing a 3:5 -internucleotidic linkage.* Journal of the Chemical Society (Resumed). 1955; 0: 2632–2638.
2. Agarwal, KL. *et al. Total synthesis of the gene for an alanine transfer ribonucleic acid from yeast.* Nature. 1997; 227: 27–34.
3. Smith HO, *et al. Generating a synthetic genome by whole genome assembly: phiX174 bacteriophage from synthetic oligonucleotides.* Proc Natl Acad Sci U S A. 2003; 100(26):15440-5.
4. Tian J. *et al. Advancing high-throughput gene synthesis technology.* Mol Biosyst. 2009; 5(7):714-22.
5. Rothstein RJ. *One-step gene disruption in yeast.* Methods Enzymol. 1983; 101:202-11.
6. Kimmel AR. And Faix J. *Generation of multiple knockout mutants using the Cre-loxP system.* Methods Mol Biol. 2006; 346:187-99.
7. Gaj, T., *et al. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering.* Trends Biotechnol. 2013; 31:397–405.
8. Yousefi, S. *et al. Catapult-like release of mitochondrial DNA by eosinophils contributes to antibacterial defense.* Nat. Med. 2008; (14) 949–953 (2008).
9. Fuchs, TA. *et al. Extracellular DNA traps promote thrombosis.* Proc. Natl. Acad. Sci. U.S.A. 2010; 10715880–5.
10. Wen, F., *et al. Extracellular DNA in pancreatic cancer promotes cell invasion and metastasis.* Cancer Res. 2013; 73: 4256–4266.
11. Gloag, ES. *et al. Self-organization of bacterial biofilms is facilitated by extracellular DNA.* Proc. Natl. Acad. Sci. U.S.A. 2013; 110: 11541–16.
12. Wang, L. and Mu, FY. *A Web-based design center for vector-based siRNA and siRNA cassette.* Bioinformatics. 2004; 20:1818–20.
13. Kahl LJ. And Endy D. *A survey of enabling technologies in synthetic biology.* J Biol Eng. 2013;7(1):13.
14. Lienert F. *et al. Synthetic biology in mammalian cells: next generation research tools and therapeutics.* Nat Rev Mol Cell Biol. 2014; 15(2):95-107.
15. Agrawal N. *et al. RNA interference: biology, mechanism, and applications.* Microbiol Mol Biol Rev. 2003; 67(4):657-85.
16. Xiong AS. *et al. Chemical gene synthesis: strategies, softwares, error corrections, and applications.* FEMS Microbiol Rev. 2008; 32(3):522-40.
17. Szybalski W. and Skalka A. *Nobel prizes and restriction enzymes*. Gene. 1978; 4(3):181-2.

18. Endy D. *Foundations for engineering biology.* Nature. 2005; 438(7067):449-53.
19. Boyle PM. and Silver PA. *Parts plus pipes: synthetic biology approaches to metabolic engineering.* Metab Eng. 2012; 14(3):223-32.
20. Kennedy CJ. *et al. Systems-level engineering of nonfermentative metabolism in yeast.* Genetics. 2009; 183:385–97.
21. Salis HM. *The ribosome binding site calculator.* Methods Enzymol. 2011; 498:19-42.
22. Pfleger BF. *et al. Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. Nature Biotechnology.* 2006; 24:1027–32.
23. Myhrvold C. and Silver PA. *Using synthetic RNAs as scaffolds and regulators.* Nature. Struct Mol Biol. 2015; 22(1):8-10.
24. Chandrasekaran AR. *Programmable DNA scaffolds for spatially-ordered protein assembly.* Nanoscale. 2016; 8(8):4436-46.
25. Yang YR. *et al. DNA Nanostructures as Programmable Biomolecular Scaffolds.* Bioconjug Chem. 2015; 26(8):1381-95.
26. Mee MT. and Wang HH. *Engineering ecosystems and synthetic ecologies.* Mol Biosyst. 2012; 8(10):2470-83.
27. Wang P, *et al. Intersectional Cre driver lines generated using split-intein mediated split-Cre reconstitution.* Sci Rep. 2012; 2:497.
28. Hirrlinger J., *et al. Split-cre complementation indicates coincident activity of different genes in vivo.* PLoS One. 2009; 4:e4286.
29. Wang HH. *et al. Programming cells by multiplex genome engineering and accelerated evolution.* Nature. 2009; 460:894–8.
30. Ma X. *et al. A. Robust CRISPR/Cas9 system for convenient, high-efficiency multiplex genome editing in monocot and dicot plants.* Mol Plant. 2015; 8(8):1274-84.
31. Lajoie MJ. *et al. Overcoming challenges in engineering the genetic code.* J Mol Biol. 2016; 428(5):1004-21.
32. Gibson DG. *et al. Chemical synthesis of the mouse mitochondrial genome.* Nat Methods. 2010; 7:901–3.
33. Hutchison CA. *et al. Design and synthesis of a minimal bacterial genome.* Science. 2016; 351(6280):aad6253.
34. Dymond JS. *et al. Synthetic chromosome arms function in yeast and generate phenotypic diversity by design.* Nature. 2011; 477(7365):471-6.
35. EMC. *The Digital Universe of Opportunities.* Infobrief. Market report analyzing the global digital data landscape. 2014:1–17.
36. O' Driscoll A. and Sleator RD. *Synthetic DNA: the next generation of big data storage. Bioengineered.* 2013 May; 4(3):123-5.
37. Zakeri B. and Lu TK. *DNA nanotechnology: new adventures for an old warhorse.* Curr Opin Chem Biol. 2015; 28:9-14.
38. Seeman NC. *Nucleic acid junctions and lattices.* J Theor Biol. 1982; 99(2):237-47.
39. Rothemund PW. *Folding DNA to create nanoscale shapes and patterns.* Nature. 2006; 440(7082):297-302.
40. Tintoré M. *et al. DNA nanoarchitectures: steps towards biological applications.* Chembiochem. 2014; 15(10):1374-90.

41. Xu GL. and Bestor TH. *Cytosine methylation targetted to pre-determined sequences.* Nat Genet. 1997; 17:376–8.
42. Maeder ML, *et al. Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins.* Nat Biotechnol. 2013; 31(12):1137-42
43. Mendenhall EM. *et al. Locus-specific editing of histone modifications at endogenous enhancers.* Nat Biotechnol. 2013; 31(12):1133-6.
44. Kang T. *et al. Reverse engineering validation using a benchmark synthetic gene circuit in human cells.* ACS Synth Biol. 2013; 2:255–62.
45. Bleris L. *et al. Synthetic incoherent feedforward circuits show adaptation to the amount of their genetic template.* Mol Syst Biol. 2011; 7:519.
46. Smith RP. *et al. Massively parallel decoding of mammalian regulatory sequences supports a flexible organizational model.* Nat Genet. 2013; 45:1021–8.
47. Pardee K. *et al. Paper-based synthetic gene networks.* Cell. 2014; 159(4):940-54.
48. Pardee K. *et al. Rapid, Low-Cost Detection of Zika Virus Using Programmable Biomolecular Components.* Cell. 2016; 165(5):1255-66.
49. Li Y. *et al. Transcription activator-like effector hybrids for conditional control and rewiring of chromosomal transgene expression.* Sci Rep. 2012; 2:897
50. Whitfield TW. *et al. Functional analysis of transcription factor binding sites in human promoters.* Genome Biol. 2012; 13:R50.
51. Huang J. *et al. CRISPR editing in biological and biomedical investigation. J Cell Physiol.* 2017; 118(12):4152-62.
52. Xiao-Jie L. *et al. CRISPR-Cas9: a new and promising player in gene therapy.* J Med Genet. 2015; 52(5):289-96.
53. Frey N. *The what, when and how of CAR T cell therapy for ALL.* Best Pract Res Clin Haematol. 2017; 30(3):275-81.
54. Julve Parreño JM. *et al. A synthetic biology approach for consistent production of plant-made recombinant polyclonal antibodies against snake venom toxins.* Plant Biotechnol J. 2017 Aug 29.
55. Parola C. *et al. Integrating high-throughput screening and sequencing for monoclonal antibody discovery and engineering.* Immunology. 2017; 153(1):31-41.
56. Ferraro B. *et al. Clinical applications of DNA vaccines: current progress.* Clin Infect Dis. 2011; 53(3):296-302.
57. Kindsmüller K. and Wagner R. *Synthetic biology: impact on the design of innovative vaccines.* Hum Vaccin. 2011; 7(6):658-62.
58. Dormitzer PR, *et al. Synthetic generation of influenza vaccine viruses for rapid response to pandemics.* Sci Transl Med. 2013; 5:185ra68.
59. Boudreau JE. *et al. Engineering dendritic cells to enhance cancer immunotherapy.* Mol Ther. 2011; 19:841–53.
60. Davies JA. and Cachat E. *Synthetic biology meets tissue engineering.* Biochem Soc Trans. 2016; 44(3):696-701.
61. Cachat E. and Davies JA. *Application of Synthetic Biology to Regenerative Medicine.* J Bioeng Biomed Sci. 2011; S2:003.
62. Graham AD. *et al. High-Resolution Patterned Cellular Constructs by Droplet-Based 3D Printing.* Sci Rep. 2017; 7(1):7004.

63. Zhou S. *Synthetic biology: Bacteria synchronized for drug delivery.* Nature. 2016; 536(7614):33-4.
64. Claesen J. and Fischbach MA. *Synthetic microbes as drug delivery systems.* ACS Synth Biol. 2015; 4(4):358-64.
65. Ye H, *et al. Pharmaceutically controlled designer circuit for the treatment of the metabolic syndrome.* Proc Natl Acad Sci U S A. 2013; 110:141–6.
66. Folcher M. and Fussenegger M. *Synthetic biology advancing clinical applications.* Curr Opin Chem Biol. 2012; 16:345–54.
67. Petters RM. *Recombinant DNA, gene transfer and the future of animal agriculture.* J Anim Sci. 1986; 62(6):1759-68.
68. Ahmad P. *et al. Role of transgenic plants in agriculture and biopharming.* Biotechnol Adv. 2012; 30(3):524-40.
69. Tyagi A. *et al. Synthetic biology: applications in the food sector.* Crit Rev Food Sci Nutr. 2016; 56(11):1777-89.
70. Tschofen M. *et al. Plant molecular farming: much more than medicines.* Annu Rev Anal Chem. 2016; 9(1):271-94.
71. Hood EE. *Plant-based biofuels.* F1000 Research 5 (2016): F1000 Faculty Rev–185.
72. Świątkiewicz S. *et al. The use of genetic engineering techniques to improve the lipid composition in meat, milk and fish products: a review.* Animal. 2015; 9(4):696-706.
73. Ravikumar S. *et al. Engineered microbial biosensors based on bacterial two-component systems as synthetic biotechnology platforms in bioremediation and biorefinery.* Microb Cell Fact. 2017; 16(1):62.
74. König H. *et al. Synthetic genomics and synthetic biology applications between hopes and concerns.* Current Genomics 14.1 (2013): 11–24.
75. Yim H. *et al. Metabolic engineering of Escherichia coli for direct production of 1,4-butanediol.* Nat. Chem. Biol. 2011; 7:445–452.
76. De Lorenzo V. *et al. Bioremediation at a global scale: from the test tube to planet earth.* Microbial Biotechnology. 2016; 9(5):618–625.
77. Caruthers MH. *et al. Deoxyoligonucleotide synthesis via the phosphoramidite method.* Gene Amplif Anal. 1983; 3:1-26.
78. Rayner S. *et al. MerMade: an oligodeoxyribonucleotide synthesizer for high throughput oligonucleotide production in dual 96-well plates.* Genome Res. 1998; 8(7):741-7.
79. Kosuri S, and Church GM. *Large-scale de novo DNA synthesis: technologies and applications.* Nat Methods. 2014; 11(5):499-507.
80. Ma S. *et al. DNA synthesis, assembly and applications in synthetic biology.* Curr Opin Chem Biol. 2012; 16(3-4):260-7.
81. Gao X. *et al. Thermodynamically balanced inside-out (TBIO) PCR-based gene synthesis: a novel method of primer design for high-fidelity assembly of longer gene sequences.* Nucleic Acids Res. 2003; 31(22):e143.
82. Hughes RA. *et al. Gene synthesis: methods and applications.* Methods Enzymol. 2011; 498:277-309.
83. Gibson DG. *et al. Enzymatic assembly of DNA molecules up to several hundred kilobases.* Nat Methods. 2009; 6(5):343-5.

84. Chalmers. F. *et al. Scaling up the ligase chain reaction-based approach to gene synthesis.* BioTechniques. 2001; 30, 249–52.
85. Li MZ. and Elledge SJ. *Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC.* Nat Methods. 2007; 4(3):251-6.
86. Czar MJ. *et al. Gene synthesis demystified.* Trends Biotechnol. 2009; 27(2):63-72.
87. Casini A. *et al. Bricks and blueprints: methods and standards for DNA assembly.* Nat Rev Mol Cell Biol. 2015; 16(9):568-76.
88. Sandhu, GS. *et al. Dual asymmetric PCR: one-step construction of synthetic genes.* BioTechniques. 1992; 12:14–16.
89. Horton, RM. *et al. Gene splicing by overlap extension.* Meth. Enzymol. 1993; 217:270–279.
90. Stemmer WP. *et al. Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides.* Gene. 1995; 164:49–53.
91. Wooddell C I. and Burgess R R. *Use of asymmetric PCR to generate long primers and single-stranded DNA for incorporating cross-linking analogs into specific sites in a DNA probe.* Genome Res. 1996; 6:886–892.
92. Young L. and Dong Q. *Two-step total gene synthesis method.* Nucleic Acids Res. 2004; 32, e59.
93. Tian, J. *et al. Accurate multiplex gene synthesis from programmable DNA microchips.* Nature. 2004; 432:1050–1054.
94. Wu, G. *et al. Simplified gene synthesis: a one-step approach to PCR-based gene construction.* J. Biotechnol. 2006; 124:496–503.
95. Kong DS. *et al. Parallel gene synthesis in a microfluidic device.* Nucleic Acids Res. 2007; 35: e61.
96. Yehezkel BT. *et al. De novo DNA synthesis using single molecule PCR.* Nucleic Acids Res. 2008; 36: e107.
97. Huang MC. *et al. TopDown real-time gene synthesis.* Methods Mol. Biol. 2012; 852:23–34.
98. Eren M. and Swenson RP. *Chemical synthesis and expression of a synthetic gene for the flavodoxin from Clostridium MP.* J. Biol. Chem.1989; 264:14874–9.
99. Mehta DV. *et al. Optimized gene synthesis, high level expression, isotopic enrichment, and refolding of human interleukin-5.* Protein Expr. Purif. 1997; 11:86–94.
100. Au LC. *et al. Gene synthesis by a LCR-based approach: high-level production of leptin-L54 using synthetic gene in Escherichia coli. Biochem.* Biophys. Res. Commun. 1998; 248:200–3.
101. Kelly, J. R. *et al. Measuring the activity of BioBrick promoters using an in vivo reference standard.* J Biol Eng. 2009; 3:4.
102. Li MZ. and Elledge SJ. *SLIC: a method for sequence- and ligation-independent cloning.* Methods Mol. Biol. 2012; 852:51–59.
103. Gibson DG. *et al. Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome.* Science. 2008; 319:1215–20.
104. Ho-Shing O. *et al. Assembly of standardized DNA parts using BioBrick ends in E. coli.* Methods Mol. Biol.2012; 852:61–76.

Notes

GenScript's Molecular Biology Service Portfolio

Gene Synthesis

Gene Synthesis
Any DNA sequence you desire!

- Guaranteed 100% sequence fidelity
- Free codon optimization

GenBrick™ Synthesis
Building blocks for synthetic biology

- Up to 200 kb long DNA
- Starting from 23 business days

Express Gene Synthesis
Speed comes from experience!

- Two flexible options: Rush or Fast
- Delivery in as few as 4 business days
- Fastest turnaround time in the market

GenPlus™ High-Throughput Gene Synthesis
Next-generation gene synthesis!

- Lowest price for high-throughput gene synthesis
- Unprecedented capacity of 100 million bp/month

Gene Synthesis Bundled Services

Bundled service options offer one-stop solutions for downstream applications; helping you save money and accelerate your research.

Express Mutagenesis Service

- Cut your turnaround time by half by bundling mutagenesis with gene synthesis
- Same unparalleled accuracy and flexibility as our Site-Directed Mutagenesis Service

Express Cloning Prep Service

- Over 150 free, popular vectors to choose from when bundled with gene synthesis
- Decrease turnaround times to as little as 2 business days when bundled with gene synthesis

Express Plasmid Prep Service

- High-quality plasmid preps with customizable quality control measures,
- Transfection grade or research grade
- Decrease turnaround times to as little as 3 business days when bundled with gene synthesis

Gene Variant Libraries

You design it, we build it!

- Gene variant libraries of any size and complexity with unprecedented efficiency
- Up to 2,000 unique DNA sequences in 30 days
- Widespread applications in:
  - Discovery biology
  - Protein engineering
  - Synthetic biology
- Variety of options:
  - Site-directed mutagenesis libraries
  - Scanning point mutation libraries
  - Randomized and degenerated libraries
  - Truncation variant libraries
  - Combinatorial assembly libraries
  - Saturation mutagenesis library
  - Combinatorial mutagenesis library

Molecular Cloning

GenEZ™ ORF Clone

Sequence-verified ORF clones!

- > 2 million ORF cDNAs
- Starting from $99/clone
- Mutant clone from $149/clone
- Custom cloning into > 100 vectors

DNA Sequencing (North America)

Rapid sequencing at low prices!

- Bulk discount pricing available
- Convenient online ordering
- Secure online data retrieval

Custom Oligos

Affordable modified/labeled oligos!

- Custom modifications and labeling
- DNA, RNA, or chimeric backbones
- Flexible synthesis scales

VectorArk & CloneArk™

Worry & error-free storage!

- Archive, share, and reuse your vectors/clones
- Free archiving of vectors when bundled with Subcloning Service
- Automatic storage of clones made by GenScript

GenParts™ Gene Fragments

A versatile tool with numerous applications!

- Fast and reliable
- Adapter-free
- Highest cloning success rate and lowest error rate in the industry
- Free codon optimization available
- Can be independently or assembled into full length genes and pathways

www.GenScript.com

GenScript USA Inc.
860 Centennial Ave.
Piscataway, NJ 08854 USA
Phone: 1-732-885-9188
Toll-Free: 1-877-436-7274
Fax: 1-732-885-5878