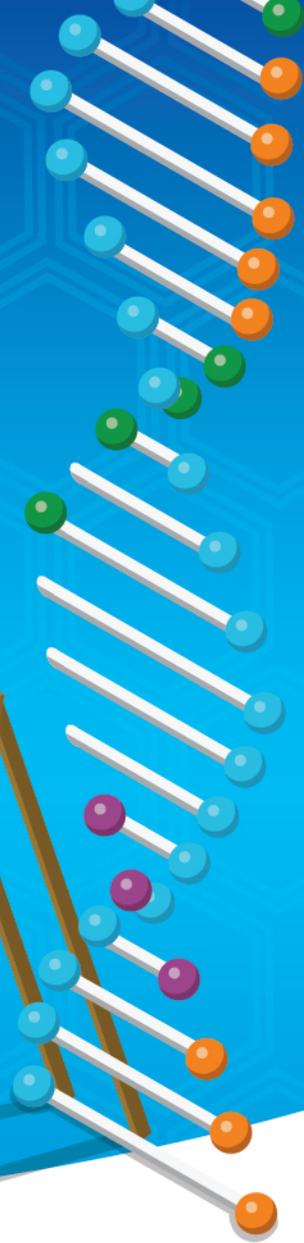


# Gene Synthesis Handbook

Third Edition

[www.GenScript.com](http://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



## Applications & Technologies for DNA synthesis

\*Updated September 2016 to include HT plasmid prep service and Express plasmid prep service

# WHAT'S ON YOUR WISHLIST?

1,563 bp - INSULIN GENE

16,569 bp - ALL HUMAN mtDNA

152,261 bp - HERPESVIRUS 1 GENOME

4,639,221 bp - E. COLI BACTERIAL GENOME

12,495,682 bp - S. CEREVISIAE YEAST GENOME

**100 MILLION BASE PAIRS**

100,258,171 bp - C. ELEGANS NEMATODE GENOME

3,095,677,412 bp - ENTIRE HOMO SAPIENS GENOME

**GenPlus™ Next-Generation technology offers the unprecedented capacity to synthesize 100,000,000 bp each month.**

Remember when you spent a month to clone a single gene? Now we can deliver 100,000 genes in that time-frame - synthesized according to your custom design, individually cloned, and sequence-perfect, guaranteed.

Download the Gene Synthesis Handbook to see how others are using this technology to advance their research.



Your on-demand genes are waiting.  
[www.genscript.com/genplus.html](http://www.genscript.com/genplus.html)

## Table of Contents

### What is Gene Synthesis?

- Overview and history of gene synthesis technology ..... 2
- What can be synthesized? ..... 4
- Gene synthesis and molecular cloning ..... 6

### Why is Gene Synthesis a revolutionary tool for biology research?

- Gene synthesis powers novel findings across research disciplines ..... 8
- Enhanced protein expression through codon optimization ..... 10
  - Case Study: Codon Optimization can improve expression of human genes in *Escherichia coli*: A multi-gene study ..... 12
  - Case Study: Codon optimization can be used to increase expression levels of drug target proteins in *E. coli* ..... 13
- Mutagenesis & Variant Libraries ..... 14
  - Case Study: High-throughput mutant gene synthesis to accelerate drug development ..... 15
- Bioengineering / Synthetic Biology ..... 16
  - Metabolic Engineering ..... 17
  - Case Study: A new way to create transgenic animal lines ..... 18
  - Case Study: The synthetic yeast genome project ..... 18

### How is Gene Synthesis performed?

- Step 1: Sequence optimization and oligo design ..... 20
- Step 2: Oligo synthesis ..... 23
- Step 3: Gene assembly ..... 26
- Step 4: Sequence verification and error correction ..... 29
- Step 5: Preparing synthetic DNA for downstream applications ..... 30
- Biosafety of synthetic genes ..... 31

### The Future of Gene Synthesis ..... 32

### References ..... 33



HT Plasmid Prep Service .....35

**Coming Soon!** Express Plasmid Prep Service .....35

# What is Gene Synthesis?

Gene synthesis refers to chemically synthesizing a strand of DNA base-by-base. Unlike DNA replication that occurs in cells or by Polymerase Chain Reaction (PCR), gene synthesis does not require a template strand. Rather, gene synthesis involves the step-wise addition of nucleotides to a single-stranded molecule, which then serves as a template for creation of a complementary strand. Gene synthesis is the fundamental technology upon which the field of synthetic biology has been built.

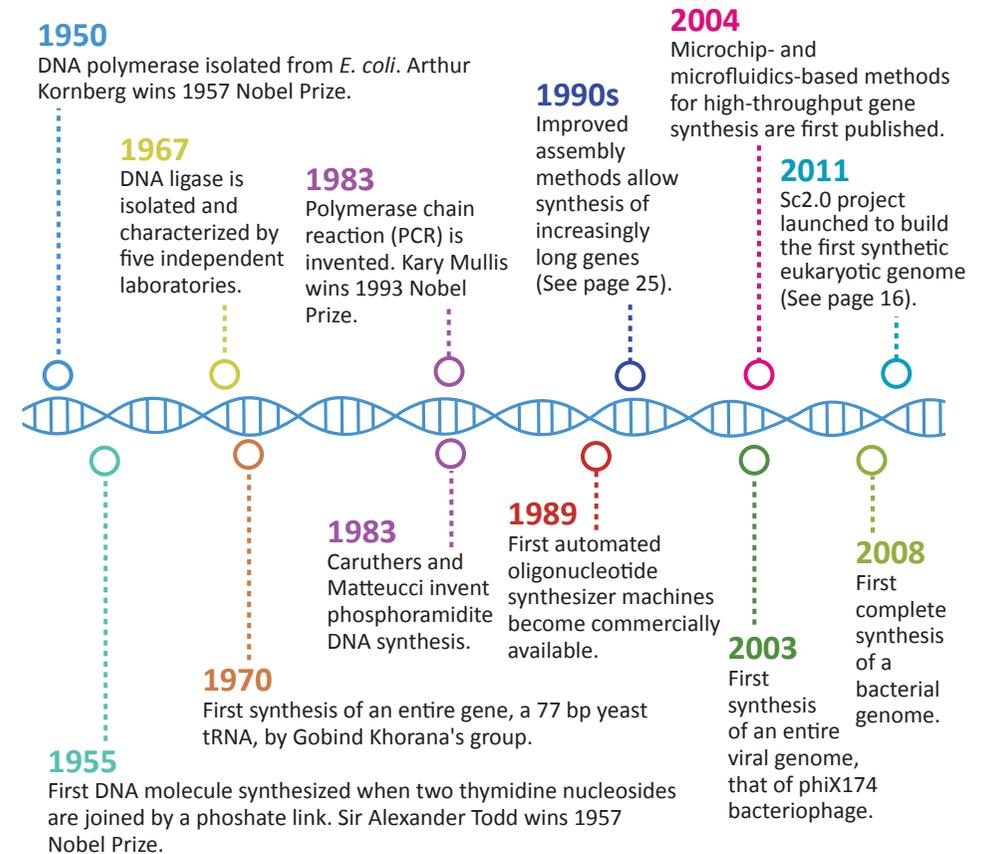
Any DNA sequence can be synthesized, including sequences that do not exist in nature, or variants on naturally-occurring sequences that would be tedious to produce through site-directed mutagenesis, such as codon-optimized sequences for increased heterologous protein expression. Synthetic DNA can be cloned into expression vectors and used in any protocol that requires natural or recombinant DNA. Synthetic genes are used to study all the diverse biological roles that nucleic acids play, from encoding proteins and regulating gene expression in the nucleus, to mediating cell-cell communication and building biofilms from extracellular DNA.

Gene synthesis technologies have revolutionized biology research. Scientists are no longer limited to classical methods of manipulating a single gene at a time; they now have the power to design or reprogram entire genomes and cells. We can quickly synthesize newly identified viral genomes to accelerate vaccine development. We can engineer novel enzymes that fight cancer and produce sustainable biofuels. We can improve crop yield and reduce vulnerability to the common pests and plant diseases that endanger food supplies and contribute to global hunger. We can detect and break down environmental pollutants in the soil, air and water. We can build designer metabolic circuits using interchangeable synthetic parts. We can create synthetic genomes and artificial cells to better understand the basic requirements of life. And life science researchers across many disciplines can do more sophisticated experiments, in less time, on smaller budgets, accelerating advances in healthcare, agriculture, energy, and other fields of human endeavor.

Methods for *de novo* chemical synthesis of DNA have been refined over the past 60 years. Synthetic short oligonucleotides (oligos) serve as custom primers and probes for a wide variety of applications. Longer sequences that serve as genes or even whole genomes can be synthesized as well; these sequences are typically produced by synthesizing 40-200 bp oligos and then assembling them in the proper order. Many methods for oligo assembly have been developed that rely upon a DNA polymerase enzyme for PCR-based amplification, a DNA ligase enzyme for ligation of oligos, or enzymes that mediate homologous recombination *in vitro* or *in vivo*. Most sequences up to 1,000 base pairs (1 kb) can be assembled in a standard molecular biology lab, and commercial gene synthesis providers routinely synthesize sequences over 10 kb.

The history of gene synthesis began in 1955, when Sir Alexander Todd published a chemical method for creating a phosphate link between two thymidine nucleosides, effectively describing the first artificial synthesis of a DNA molecule.<sup>1</sup> The first successful synthesis of an entire gene was reported by Gobind Khorana's group in 1970; the 77 bp DNA fragment took 5 years to synthesize.<sup>2</sup> Subsequent improvements in DNA synthesis, sequencing, amplification, and automation have made it possible now to synthesize genes over 1 kb in just a few days, and to synthesize much longer sequences including entire genomes. Gene synthesis can now be easily and cost-effectively outsourced to commercial providers. **GenScript, a pioneer in gene synthesis, was founded in 2002 and is the largest gene synthesis supplier in the world.**

Figure 1: Timeline of Gene Synthesis Technology Development



## What can be synthesized?

Gene synthesis can generate recombinant, mutated, or completely novel DNA sequences without a template, simplifying the creation of DNA tools that would be laborious to produce through traditional molecular cloning techniques. A wide variety of types of sequences can be produced to aid in diverse research applications (See table 1). In addition to DNA sequences, RNA and oligos containing modified bases or chimeric DNA-RNA backbones can also be synthesized. However, the most widely used synthetic sequences are customized DNA of the following types:

**cDNA** corresponding to the **open reading frames (ORFs)** of genes can be synthesized for overexpression or heterologous expression. Synthesized genes can encode either naturally occurring transcripts or custom-designed variants such as point-mutants or fluorescently tagged reporters. Gene synthesis accelerates the creation of **recombinant coding sequences** such as fusion proteins or **promoter-reporter constructs** compared to traditional molecular subcloning and mutagenesis techniques. Protein expression levels can be greatly improved through codon optimization (See page 10), which is simple to perform as part of synthetic gene design but practically impossible to perform using traditional recombinant DNA technology.

**Genomic DNA** can be synthesized for the study of gene regulatory elements, mRNA processing, or to create artificial genes or genomes. Synthetic genomic DNA can be studied *in vitro*, inserted into host cells in culture, or used to create transgenic animal lines. Conventional gene targeting of embryonic stem cells with disruption cassettes inserted through restriction enzymes<sup>3</sup> or the Cre-loxp system<sup>4</sup> can be made simpler, more rapid, and less expensive 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 Effector Nucleases (TALENs).<sup>5</sup>

**RNA interference (RNAi)** technology can be used to suppress gene expression. Small interfering RNA (siRNA) sequences of 21-25 bp block gene expression by targeting complimentary sequences within mRNA for degradation or sequestration. siRNA can be synthesized as RNA oligonucleotides annealed to form an RNA duplex, but these molecules are short-lived and often difficult to deliver into cells. Vector-based DNA constructs that encode short hairpin RNA (shRNA) are easier to deliver through viral transduction or plasmid transfection or electroporation, and can stably express 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. siRNA-mediated gene knock-down enables

the study of genes whose complete knock-out would be lethal, and when combined with lentiviral delivery, vector-based siRNA can cost-effectively generate stable cell lines or transgenic animals.

*GenScript's OptimumGene™ codong optimization platform can help you optimize expression of your gene in any host through the use of a proprietary PSO algorithm that takes into account multiple variables during the optimization process.*

**Extracellular DNA** has recently been shown to play important roles in innate immune responses<sup>7</sup>, thrombosis<sup>8</sup>, cancer metastasis<sup>9</sup>, biofilm formation<sup>10</sup> and other biological functions. Although we normally think of DNA as simply the blueprint specifying the construction of enzymes which do the work of living cells, nucleic acids are in fact versatile biomolecules whose many roles are still being elucidated. Gene synthesis technology can fabricate any DNA sequence to study its roles both inside and outside of the cell.

**Table 1: Types of Synthetic DNA and their Research Applications**

Types of Synthetic DNA	Research Applications
cDNA / ORFs	Over-expression or heterologous expression
Customized coding Sequences	Expressing fusion proteins; high-level protein expression from codon-optimized sequences for purification for 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; studying gene structure, regulation, and evolution.
Mutant sequences	Confirming function (promoter-bashing; amino acid substitutions); Protein engineering (rational design or screening/directed evolution)
RNAi constructs (shRNA, miRNA, siRNA)	Regulating (suppressing) gene expression; intercellular communication
Extracellular DNA	Biofilm formation, intercellular signaling as in cancer metastasis

## Gene Synthesis and Molecular Cloning

The *de novo* chemical synthesis of DNA differs from traditional molecular cloning in that it does not require a template. Gene synthesis allows researchers to specify a desired sequence and custom-build it directly. Gene synthesis is frequently far more straightforward, faster, and less costly than using alternative methods of molecular cloning and mutagenesis.

Researchers experienced with molecular cloning know that, despite improvements over the past several decades in recombinant DNA tools tools, such as enzymes and cloning vectors, getting the clone you want is hardly a fool-proof endeavor. Even the seemingly simple task of isolating a gene using PCR cloning or restriction digestion can be tedious and error-prone depending on the sequence and its context in the source material. Numerous cDNAs predicted by genomic or partial cDNA information have proven difficult to clone but simple to synthesize. Recombining several gene cassettes to form fusion proteins can create even more challenges in selecting unique cutters, performing iterative digests and ligations, and ensuring that all assembled portions maintain the correct open reading frame.

Once molecular cloning is successful, constructs designed for heterologous expression often yield protein levels so low that it becomes difficult to do the experiment for which the construct was designed. Codon optimization can overcome this challenge by modifying a sequence to promote efficient heterologous expression (for more details, see page 10). Codon optimization can be performed prior to *de novo* synthesis to yield higher levels of protein expression and facilitate the study of function; this is not possible with traditional molecular cloning which preserves a template found in nature.

Gene synthesis can be combined with molecular cloning to make the synthetic products ready to use in research applications. Placing synthetic genes into a plasmid vector offers several advantages: it protects the DNA against degradation during storage; it allows complete sequence verification using primer sites on the vector that flank the gene insert; it facilitates clonal amplification in a transformation-competent bacterial host; and it facilitates shuttling of the synthetic gene into expression vectors for transfection or electroporation into host cells of interest for transient or stable expression.

Bare DNA fragments can be used directly in certain applications. For example, 10-40 bp oligos are used as primers in PCR. Longer sequences may be used as templates for radiolabelled probe synthesis for *in situ* hybridization or other techniques. *In vitro* transcription and translation protocols employ bare DNA fragments as well.

*GenEZ™ Next-Generation Molecular Cloning from GenScript eliminates the hands-on time and uncertainty of traditional molecular cloning: simply enter your desired insert DNA sequence online, select your vector, and receive your sequence-perfect clone with guaranteed on-time delivery.*

**Table 2: Comparing Gene Synthesis and Molecular Cloning**

Technique	Features
Traditional Molecular Cloning	<ul style="list-style-type: none"><li>• limited to sequences that appear in nature</li><li>• codon optimization is not feasible</li><li>• time- and resource-intensive for end-user</li></ul>
Gene Synthesis of Bare DNA Fragments	<ul style="list-style-type: none"><li>• customizable sequence—no template required</li><li>• error-prone; yields mixed pools of variants</li><li>• time consuming to verify sequences</li><li>• lower cost for “quick and dirty” high-content, high-throughput screens</li></ul>
Gene Synthesis with Subcloning into a Plasmid Vector	<ul style="list-style-type: none"><li>• customizable sequence—no template required</li><li>• uniformly accurate sequences from a clone</li><li>• easy to verify complete sequence using primers that flank gene insert</li><li>• fast and cost-effective</li></ul>

*GenScript delivers all gene synthesis products inserted into the plasmid vector of your choice according to your preferred cloning strategy. Sequence chromatograms covering the entire gene insert provide assurance of 100% sequence accuracy for every gene synthesis order.*

# Why is Gene Synthesis a revolutionary tool for biology research?

Gene Synthesis forms the foundation of the new field of synthetic biology. Gene synthesis is also accelerating research in well-established research fields by providing critical advantages over more laborious traditional molecular cloning techniques. *De novo* gene synthesis is required when template DNA molecules are not available, such as for codon-optimized sequences. In addition, it is frequently a cost-effective alternative to traditional molecular cloning techniques, such as when multiple DNA segments are being recombined. The case studies below illustrate the power of customizable design afforded by gene synthesis.

**Table 3: Gene Synthesis Powers Novel Findings Across Research Disciplines**

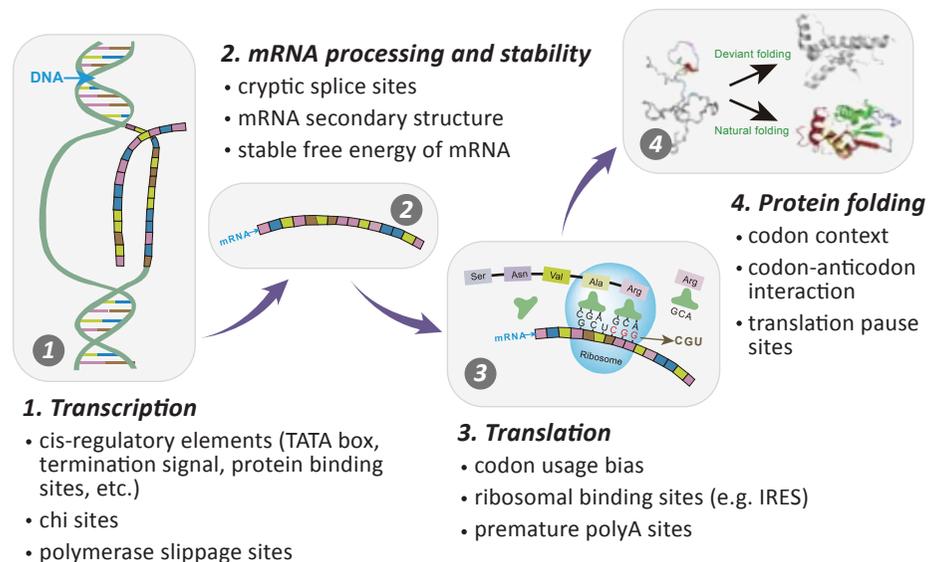
<p><b>Cancer</b></p>	<p><b>Characterization of Dicer1e protein reveals important role in oral squamous cell carcinomas.</b><sup>11</sup> Cantini et al., using a synthesized Dicer1e gene, to analyze its roll in oral squamous cell carcinoma pathogenesis. Researchers found that Dicer1e protein levels were overexpressed ion cell lines with oral squamous cell carcinoma and that silencing of the gene itself was able to effectively limit cancer cell proliferation. The silencing of Dicer1e induced cell cycle arrest and apoptosis. The knockdown of Dicer1e also increased the effectiveness of chemotherapy treatment, more specifically the effectiveness of cisplatin.</p>
<p><b>Structural Biology</b></p>	<p><b>Scientists characterize atomic detailed information of the HIV-1 glycan shield, could lead to effective vaccine.</b><sup>12</sup> Stewart-Jones et al. report the atomic-level details of the glycan shields surrounding the HIV-1 Env trimers from clades A, B, and G. By rapidly synthesizing HIV-1 Env trimer designs, lead candidates that could be crystallized were identified. This lead to the eventual crystallization of the HIV-1 Env trimer and surrounding glycan shields. It is through this crystallization and characterization of the glycan shields that researchers have been able to further understand mechanisms by which HIV avoids antibody detection and subsequent attack by the immune system. Antibodies can now be designed with the glycan shield in mind, further increasing their effectiveness towards HIV-1</p>

<p><b>Vaccines/ Virology</b></p>	<p><b>Vaccine-induced antiodies that can now neutralize multiple Influenza A viruses.</b><sup>13</sup> Joyce et al. have discovered antibodies that are capable of neutralizing mulitple forms of Influenza A from subjects enrolled in an H5N1 DNA/MIV-prime-boost influenza vaccine trial. The process of validating the neutralizing antibodies found in the paper required a large production of antibodies. Gene synthesis and codon optimization allowed researchers to quickly test the antibodies in the paper for their ability to target the stem region of the virus, more specifically the protein hemagglutinin, recognizing overlapping epitopes in the hemagglutinin stem. This high conservation of the stem seen across influeza virus types results in the ability of the antibodies to recognize multiple strains of influenza, which could ultimately lead to the development of a universal influenza vaccine.</p>
<p><b>Synthetic Biology</b></p>	<p><b>Creation of minimal cell with just the genes needed for independent life.</b><sup>14</sup> Hutchison et al., building on previous research in 2010 in which a self-replicating, synthetic bacterial cell was synthesized, has now been able to design and synthesize a minimal bacterial genome. This is a groundbreaking achievement in that it shows that genomes can be designed on a computer, synthesized in a lab, and then moved into a recipient bacterial cell to produce a cell that is now controlled by the synthetic genome. This synthetic bacterial cell will help provide clues and ideas into how to investigate the core functions that are required for life in living organisms.</p>
<p><b>Plant Biology</b></p>	<p><b>LqhIT2, an insect-specific toxin, is not only toxic on its own to pests, but also activates the phenylpropanoid pathway in transgenic rice, increasing insect resistance.</b><sup>15</sup> Tianpei et al., using codon optimized and synthesized LqhIT2 scorpion toxin peptide genes, found that LqhIT2 was not only toxic to insects and pests by itself, but was also able to improve insect resistance in transgenic rice by activating inherent pathways in rice. This activation occurs via elevated expression of lipoxygenase, which is a known key component of the jasmonic acid biosynthetic pathway. Several other key components are also up-regulated. All of this resulting in jasmonate-mediated priming, which increases flavonoid and lignin content, resulting in improved insect resistance in the transgenic rice tested.</p>

## Codon Optimization to Increase Protein Expression

In order to characterize a gene's function or purify a protein for structural study, researchers frequently need to express a gene in a host cell or model organism different from the one in which it appears in nature. Unfortunately, heterologously expressed genes often suffer from low rates of protein expression. A common cause of low expression levels is the variation in codon usage frequency between different species; this can cause the translation of heterologously expressed genes to stall due to tRNA scarcity, leading to lower protein levels and increased rates of improper protein folding. The degeneracy of the genetic code makes it possible to change the DNA sequence in a way that does not alter the final amino acid sequence, but does have significant effects on the efficiency of transcription, mRNA stability, translation, and protein folding. Aside from codon usage bias, many other features of a DNA sequence affect the efficiency of transcription, the proper splicing and processing of mRNA, and mRNA stability, all of which in turn reduce the ultimate protein yield (Figure 2).

**Figure 2: DNA Sequence Features that Influence Protein Expression Levels**



Modifying a DNA sequence to optimize it for efficient heterologous protein expression while preserving the amino acid sequence is called **codon optimization**. In a study by Burgess-Brown *et al.*<sup>16</sup> that investigated the expression of 30 human genes in *E. coli*, codon optimization increased both the total protein yield and the protein solubility, compared to heterologous expression of native gene sequences. Codon optimization algorithms have improved since this 2008 study, and codon optimization is now widely used by researchers to facilitate the purification of protein for structural studies, enzyme kinetics, and other biochemical investigations.

*GenScript's OptimumGene™ PSO-driven algorithm, patented in 2012, efficiently optimizes sequences by taking into account multiple 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 that frequently only consider codon usage tables when optimizing genes for expression.*

Functional studies of cell signaling can also benefit from codon optimization. For example, When Yoo *et al.* set out to investigate the physiological role of Lyn, a SRC family kinase, in a zebrafish model of wound healing, they needed to express the Lyn protein in a tractable *in vitro* cell-based model system for reductionist functional studies. They found that the native zebrafish Lyn sequence was poorly expressed in the HEK293 human cell line, so they turned to GenScript to synthesize a humanized, codon-optimized Lyn sequence. This synthetic construct yielded sufficient heterologous protein expression levels to allow characterization of the Lyn protein. Subsequent experiments identified Lyn as a redox sensor that initiates neutrophil recruitment to wounds. This finding was published in *Nature* and is exciting not only for revealing basic mechanisms of wound healing but also for providing insight on immunotherapeutic strategies for targeting tumors, which are a type of 'unhealed wound' that release H<sub>2</sub>O<sub>2</sub> and thus may be subject to regulation by Lyn.<sup>17</sup>

High-level protein expression is important not only for basic research but also for translational research and drug development, for example, by enhancing the immunogenic response of DNA vaccines. By overcoming the problem of insufficient heterologous expression levels, codon optimization serves as a valuable tool for life science research.

Codon optimization exemplifies the flexibility afforded by gene synthesis. While it is possible to overcome the limits of codon usage bias by expressing the tRNAs that correspond to rare codons, this approach does not address other optimizable features of the gene sequence, and it introduces the confound of altering the translational machinery that regulates all endogenous protein turnover in the cell. It is also possible to achieve codon optimization through site-directed mutagenesis, but given the high number of nucleotide alterations typically desired, this technique is extremely laborious and costly compared to *de novo* gene synthesis.

## Case Study: Codon Optimization can improve expression of human genes in *Escherichia coli*: A multi-gene study

**Challenge:** Increase expression of several human genes in a different host, in this case *E. coli*, for functional studies.

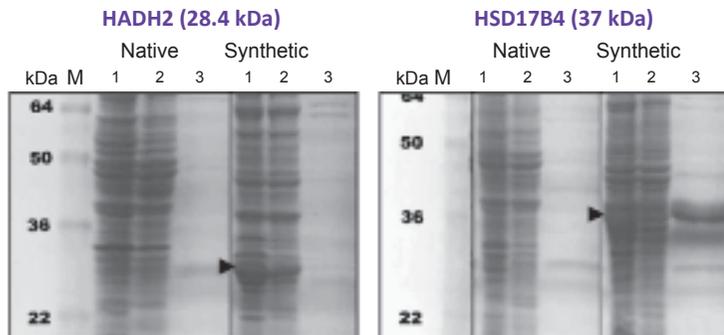
**Solution:** Synthesize genes using OptimumGene™ codon optimization technology to increase expression in a different host.

Gene Name	Native			Synthetic			Expression Syn vs Nat	Solubility Syn vs Nat
	1	2	3	1	2	3		
CBR1	■	■	■	■	■	■	▲	▲
CBR3	■	■	■	■	■	■	▲	▲
GMDS	■	■	■	■	■	■	▲	▲
HADH2	■	■	■	■	■	■	▲	▲
HSD17B2	■	■	■	■	■	■	▲	▲
HSD17B4	■	■	■	■	■	■	▲	▲
MGC4172	■	■	■	■	■	■	▲	▲
PECR	■	■	■	■	■	■	▲	▲
RETSR2	■	■	■	■	■	■	▲	▲
SPR	■	■	■	■	■	■		

■ Expressed  
 ■ Expressed, Soluble and Purified  
 ■ Not Expressed

▲ Improvement of expression and/or solubility with synthetic gene after codon optimization

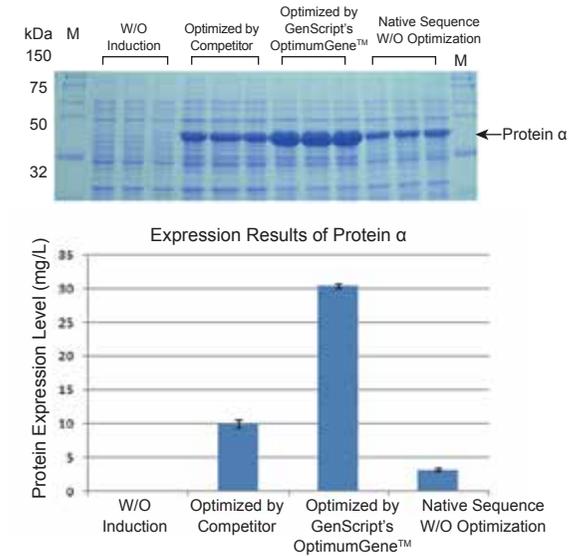
1. Total Cell Protein
2. Soluble Fraction
3. Eluted Fraction



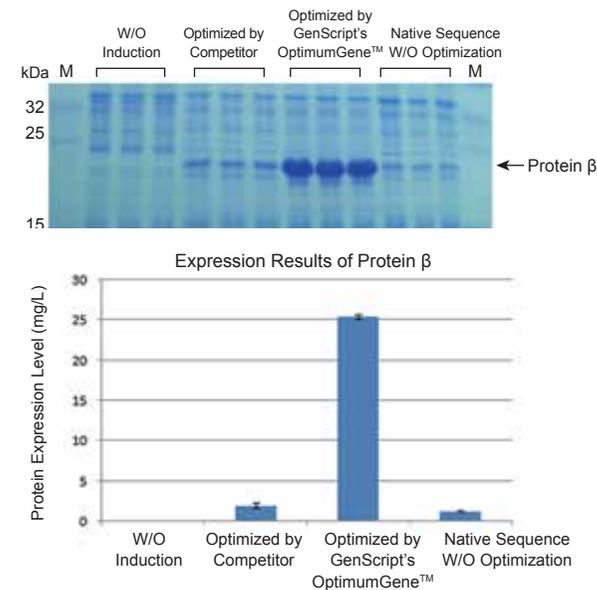
**Figure 3.** GenScript OptimumGene™ codon optimized genes increased the yield of expression (8 out of 10 genes) and the degree of solubility in some cases (6 out of 10 genes), when compared to the native genes.

## Case Study: Codon optimization can be used to increase expression levels of drug target proteins in *E. coli*

The figures below demonstrate the effectiveness of codon optimization with OptimumGene™ technology in increasing expression levels for drug target proteins in two separate cases.



**Figure 4.** GenScript's OptimumGene™ codon optimization delivered 10 times higher expression level for protein alpha when compared to the non-optimized native gene sequence. The expression level was 3 times more than that from Competitors GA's optimization method.



**Figure 5.** GenScript's OptimumGene™ codon optimization delivered 20 times higher expression level for protein beta when compared to the non-optimized native gene sequence. The expression level was 13 times more than that from Competitors GA's optimization method.

## Mutagenesis & Variant Libraries

Gene Synthesis facilitates the discovery or design of new protein variants (mutants) with properties desired for medical, industrial, agricultural, or other applications.

Traditional methods for inducing random mutations include subjecting live cells or animals to UV radiation or chemical mutagens in order to induce DNA damage. DNA repair enzymes can, in some cases, introduce or preserve mutations resulting from damage such as single- or double strand breaks, but this method leads to high rates of lethality and low rates of useful mutant phenotypes. Depending on the type of mutagen used, mutations may be biased, limiting the range of possible mutants to only a fraction of those that are theoretically possible or that can be obtained through other strategies.

Site-directed mutagenesis may be designed based on experimental data such as X-ray crystallography data that identifies residues of interest. Site-directed mutagenesis can be performed using PCR-based methods with mutated primer sequences to create and selectively amplify mutated sequences. This type of mutagenesis must be followed by sequence verification, and can be tedious, error-prone, and costly compared to *de novo* chemical synthesis of the desired mutant sequences.

DNA libraries containing many mutant sequences can be used for unbiased screens to identify novel sequences with desired structural or functional properties. If researchers have a functional assay for a desired protein activity, they can screen a mutant library, sequence the best-performing candidates, and use these sequences to produce new mutant libraries for iterative rounds of screening and improvement. This strategy, called directed evolution, allows researchers to optimize a protein's function even without understanding its mechanistic basis.<sup>18</sup> The mutagenesis strategy and technical method selected critically determine the likelihood of obtaining mutants with the desired properties. Well-prepared libraries can be engineered to contain unbiased point mutations, to maintain the open reading frame when insertions or deletions are present, and to avoid truncations.

*GenScript can deliver sequence-guaranteed mutants of up to 1 kb in length within two weeks. Using standard techniques for site-directed mutagenesis, sequence verification, and error correction in a typical biology research lab can take more than twice as long.*

Gene synthesis can create small or large numbers of systematic or random mutants more efficiently than traditional mutagenesis methods. Carefully designed mutant sequences or libraries can be produced with greater accuracy through gene synthesis than through other methods.

*GenScript offers custom synthesis of Gene Variant Libraries for any research application, such as Combinatorial Assembly Libraries for synthetic biology, Truncation Variant Libraries for structural biology, and Scanning Point Mutation Libraries for protein engineering.*

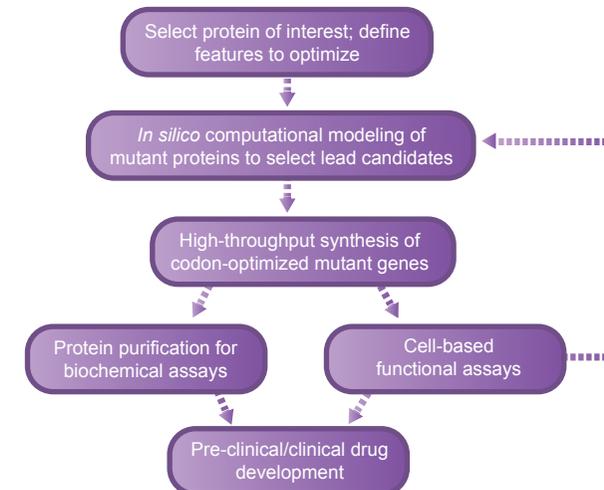
## Case Study: High-Throughput Mutant Gene Synthesis to Accelerate Drug Development

**Challenge:** Speed up R&D process for developing enzyme-based therapeutics.

**Solution:** Synthesize many mutant sequences for rapid screening.

A small company was founded by scientists whose goal is to engineer novel biologic drugs, such as antibodies and growth factors for oncology and immunology, enzyme replacement therapies for genetic disorders, and myriad other therapeutic proteins.<sup>19</sup> The company's expertise is in computational modeling of protein dynamics to design new or improved enzymes. Their virtual simulations allow them to predict which mutations will enhance protein stability, substrate affinity, enzymatic activity, or other key features that contribute to drug efficacy.

**Figure 6: Process for Engineering and Validating Novel Therapeutic Proteins**



To validate their *in silico* mutagenesis, they need to synthesize the mutated sequences, express the proteins for biochemical characterization, and perform cell-based functional assays. Promising candidates can then be licensed to pharmaceutical companies that have the infrastructure to follow through with pre-clinical and clinical trials, bringing new drugs to patients in need. Gene synthesis allows high-throughput, fast, accurate construction of mutant sequences that are then cloned into expression vectors. Gene synthesis also facilitates high-level protein expression through host cell-specific codon optimization, allowing rapid purification of sufficient quantities of protein for biochemical characterization. ***By outsourcing to GenScript, the industry leader in capacity and total automation of gene synthesis, this biotech client was able to speed up their R&D process, meet the deadlines from their pharmaceutical partners, and secure more funding from venture capital investors.***

## Synthetic Biology / Bioengineering

Bioengineers view genes, and even entire metabolic circuits in cells, as machines made up of parts that play standardized functional roles. Once these parts are identified, they can be treated as interchangeable, and rearranged in any way to create new and improved machines. While traditional molecular cloning techniques allow the creation of knock-in genes and fusion proteins (after several steps of digestion, PCR amplification, assembly, and verification), *de novo* gene synthesis streamlines the production of new biomolecular machines with customized sequences.



**Synthetic Biology**  
based on standard parts

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

## Metabolic Engineering

Metabolic engineering enables the efficient production of naturally occurring compounds that have valuable research, clinical or industrial uses. Before gene synthesis technologies existed, natural products were typically obtained either from plant extracts or from total chemical synthesis of the desired compound. In the case of some compounds, purification from natural sources can be problematic due to the need to resolve complex mixtures of closely related compounds, and our current best methods for chemical synthesis are limited by low yield and low specificity requiring additional painstaking purification. As an alternative approach, metabolic engineering allows biosynthetic pathways to be reconstructed in model organisms (typically microbiological hosts such as *Escherichia coli*) so that useful quantities of the desired product can be harvested.

In one example of metabolic engineering to improve synthesis of a desired biomolecule, Brazier-Hicks and Edwards developed a method for efficient production of C-glycosylated flavanoids for dietary studies by using gene synthesis to re-engineer a metabolic circuit in yeast.<sup>20</sup> They designed synthetic variants of five genes that comprise the flavone-C-glycoside pathway in rice plants, which were subsequently codon-optimized for expression in yeast. These synthetic genes were used to construct a polyprotein cassette that expresses the entire metabolic circuit in a single step.

Metabolic engineering has also been used to create new methods for alternative energy from unconventional sources. Dellomonaco *et al.* showed that fatty acids can serve as biomass for sustainably produced biofuels by using gene synthesis to engineer several native and heterologous fermentative pathways to function in *E.coli* under aerobic conditions.<sup>21</sup> These synthetically engineered bacteria convert fatty acid-rich feedstocks into desirable biofuels (ethanol and butanol) and biochemicals (acetate, acetone, isopropanol, succinate, and propionate), with higher yield than more widely used lignocellulosic sugars.

## Case Study: A New Way to Create Transgenic Animal Lines

**Challenge:** Expand the power of CRISPR/Cas9 nuclease-based genome editing to produce heritable changes in *Caenorhabditis elegans*.

**Solution:** Synthesize novel Cas9 and sgRNA vectors that can be used to efficiently create new transgenic strains.

Transgenic animal models serve as critical tools for many arenas of life science and medical research. CRISPR/Cas9 nuclease-based genome editing is the most recent addition to the repertoire of methods for genetic modification, which also includes TALEN, ZFN, and Cre-lox technologies. Friedland *et al.* sought to optimize the Cas9 gene so it could be used to produce heritable changes in *C. elegans*.<sup>22</sup>

Gene synthesis allowed Friedland and colleagues to boost heterologous expression of the bacterial Cas9 in a *C. elegans* host through codon optimization. It also enabled fast, error-free recombination of the multitude of critical sequences: plasmid vectors to allow efficient delivery into cells, promoters to drive efficient transcription, a nuclear localization sequence fused to the end of the Cas9 sequence to allow it to access the host's nuclear DNA, the sgRNA scaffold to allow precise targeting, and the target sequence they wished to introduce.

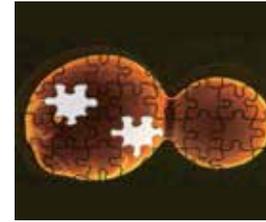
These optimized synthetic vectors can be used by other researchers in the future to efficiently bioengineer their own transgenic worm strains, simply by altering the target and sgRNA scaffold sequence to correspond to their gene of interest. The target sequence can include any mutation, insertion, or deletion the researcher desires. **GenScript can synthesize Cas9 and sgRNA sequences to enable efficient genome editing in any host.**

## Case Study: The Synthetic Yeast Genome Project

**Challenge:** Create a functional synthetic genome to support a living yeast cell.

**Solution:** Synthesis of 10 kb chunks for assembly into entire chromosomes.

The Sc2.0 Project, led by Dr. Jef Boeke at the Johns Hopkins University, is the first attempt to synthesize a eukaryotic cell genome, that of *Saccharomyces cerevisiae*. The goal of the Sc2.0 Project is to synthesize the entire yeast genome—about 6,000

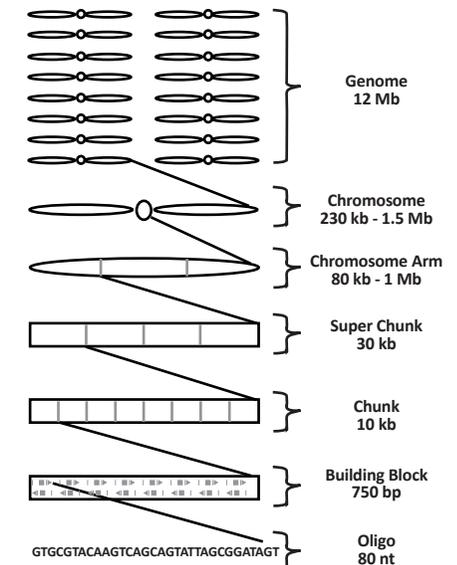


genes— with a built-in diversity generator that will enable researchers to discover how yeast, as a model organism, deals with genetic change and how genomes might be improved to create more robust organisms. This project lays the foundation for future work to design genomes for specific purposes, such as creating new medications or biofuels.

To facilitate gene rearrangement and genome minimization, Boeke and colleagues designed the Synthetic Chromosome Recombination and Modification by LoxP-mediated Evolution (SCRaMbLE) system.<sup>23</sup> More than 5,000 *loxP* sites will be introduced in the final genome, so that pulsatile Cre expression can create innumerable unique deletions and rearrangements. Viable genomes can be sequenced and many variant yeast strains can be studied to deduce the limits of chromosome structure, minimal eukaryotic genome structures, and gene/feature adjacency rules in genomes.

In 2011, the first designer yeast chromosome arm, synIXR, was successfully synthesized and shown to support robust yeast cell growth.<sup>24</sup> To construct the remainder of the synthetic genome, incorporating SCRaMbLE and other designer features, requires DNA synthesis on an unprecedented scale. To this end, in 2012 GenScript synthesized 17 DNA fragments, each ~10 kb in length to become part of synthetic yeast chromosome VI. **GenScript is proud to be a contributing partner in the Sc2.0 International Consortium whose goal is to build a designer synthetic eukaryotic genome.**

**Figure 7: Building a synthetic genome through assembly of smaller parts**



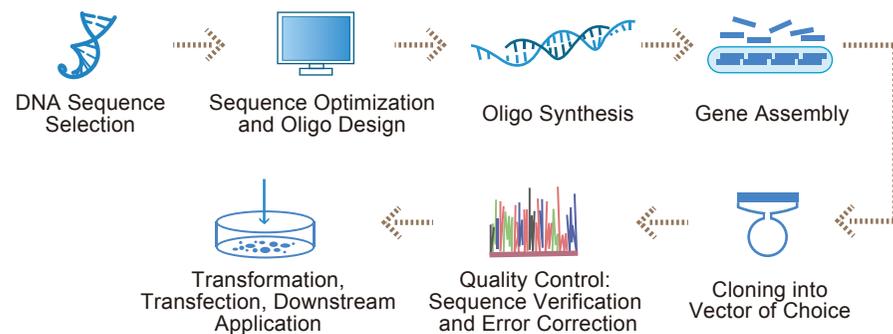
Reprinted, with permission, from Dymond, J. & Boeke, J. The *Saccharomyces cerevisiae* SCRaMbLE system and genome minimization. *Bioengineered Bugs* 3, 5–4 (2012).

# How is Gene Synthesis Performed?

## The basic steps of gene synthesis are:

1. sequence optimization and oligo design
2. oligo synthesis
3. gene assembly
4. sequence verification and error correction
5. preparing synthetic DNA for downstream applications

Figure 8: Steps in Gene Synthesis



## Step 1: Sequence Optimization and Oligo Design

### Sequence Optimization

Once you have selected your gene of interest, you must design the sequence to be synthesized. Keep the end application in mind: for example, codon optimization is appropriate if your goal is to maximize heterologous protein expression levels, but it may not be appropriate if you want to study endogenous regulation of gene expression. For constructs containing multiple segments, make sure your intended reading frame is maintained throughout the entire coding region. Frequently you'll want to add short flanking sequences to facilitate later excision or recombination, via restriction enzymes or similar tools. Be sure your sequence does not include restriction enzyme recognition sites or other sequences that might interfere with your downstream workflow. Be aware of the presence of functional domains such

as *cis*-regulatory elements or RNase splice sites, which may be unintentionally introduced during codon optimization. The presence of biologically functional sequences such as *cis*-regulatory elements or RNase splice site on oligos may hinder *in vivo* assembly, maintenance, or expression.

*GenScript's OptimumGene™ PSO-driven algorithm efficiently optimizes sequences by taking into account multiple expression variables, 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 that only consider codon usage tables.*

Some sequence features make synthesis more challenging, including: extremely high or low GC content; highly repetitive sequences; complex secondary structures such as hairpins; unstable structural elements; polyA stretches; and longer sequences, especially over 1 kb. These are all important features to consider when selecting and optimizing sequences to synthesize, and they will increase the cost and turnaround time of synthesis, whether it is performed in-house or outsourced to a commercial service provider.

### Oligo Design

After finalizing the sequence that will be synthesized, sequence analysis is required to determine the best way to divide the whole gene into fragments that will be synthesized and then assembled. For very large synthetic genes, you will typically want to divide the complete sequence into chunks of 500-1,000 kb to be synthesized separately and assembled later. Numerous oligo design software programs are available to assist in oligo design (See table 4).

While synthesis platforms utilizing phosphoramidite chemistry have very low error rates, errors do accumulate as strand length increases; therefore, gene synthesis typically uses oligos 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 the researcher's preferences. As a general rule, shorter oligos 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. Oligos should be adjusted for even lengths and equal melting temperatures. Each oligo design tool has its own default parameters; for example, Gene design uses default settings of 60 bp oligos with 20 bp overlaps, values which typically work well with yeast and mammalian sequences which are ~40% GC.

**Table 4: Oligo Design Tools**

Design Tool	Advantages	Limitations
<b>DNA Works<sup>25</sup></b> <a href="http://helixweb.nih.gov/dnaworks/">http://helixweb.nih.gov/dnaworks/</a>	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. <sup>26</sup>
<b>Gene2Oligo<sup>27</sup></b> <a href="http://berry.engin.umich.edu/gene2oligo/">http://berry.engin.umich.edu/gene2oligo/</a>	Simple user interface, designs oligos for both LCR and PCR-based assembly.	Limited to genes less than 1 kb in length.
<b>GeneDesign<sup>28</sup></b> <a href="http://baderlab.me.jhu.edu/gd/">http://baderlab.me.jhu.edu/gd/</a>	Breaks multi-kb sequences into ~500 bp blocks for initial assembly in separate pools.	Does not offer mishybridization analysis.
<b>TMPrime<sup>29</sup></b> <a href="http://prime.ibn.a-star.edu.sg">http://prime.ibn.a-star.edu.sg</a>	Designs oligos for both LCR- and PCR-based assembly. Produces the most homologous melting temperatures ( $\Delta T_m < 3^\circ\text{C}$ ) and widest range of annealing temperatures (50–70°C).	Requires thorough understanding of many input parameters for correct submission; less user-friendly interface.

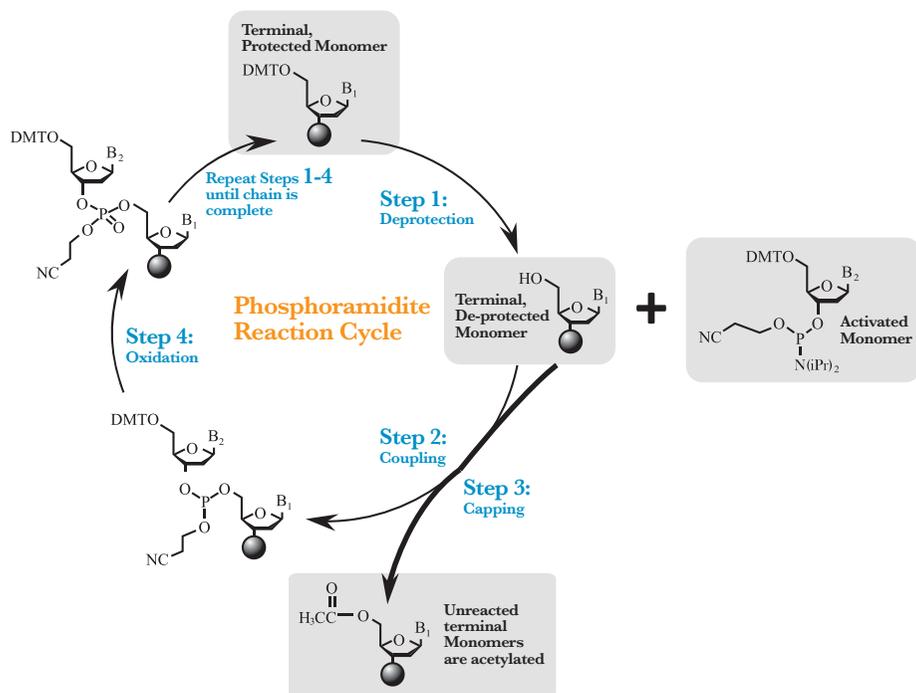
In addition to oligo length, some factors to consider in selecting oligo sequences include GC content, sequence repeats, and the tendency for hairpin formation. GC content determines the stability of DNA strands and thus the melting temperature. All *in vitro* assembly methods, both ligase and polymerase based, rely upon the melting and annealing of oligos and thus require that oligos have equal melting temperatures. Repeated sequences may pose challenges for correct annealing or ligation of oligos during assembly. Repeats may occur in many forms, including direct, inverted, palindromic, or tandem repeats. Improper hybridization or intramolecular binding such as hairpin structure formation can be avoided through careful design of oligo sequences, including the overhangs or primers used to assemble them.

*GenScript's proprietary DSOpt™ Sequence Analysis tool, developed through years of experience as the volume leader in gene synthesis, takes the guesswork out of selecting the best assembly method and oligo design strategy for each project.*

## Step 2: Oligo Synthesis

All DNA fabrication today begins with the step-wise addition of nucleotide monomers via phosphoramidite chemistry to form short oligonucleotides. Oligo synthesis via phosphoramidite chemistry uses modified nucleotides, called phosphoramidites, to ensure that nucleotides assemble in the correct way and to prevent the growing strand from engaging in undesired reactions during synthesis. The phosphoramidite group is attached to the 3' O and contains both a methylated phosphite and a protective di-isopropylamine to prevent unwanted branching. Phosphite is used because it reacts faster than phosphate. Methyl groups are attached to the phosphite, and amino-protecting groups are added to the bases, to protect against unwanted reactions until oligonucleotide synthesis is complete. Although DNA synthesis in living cells always occurs in the 5' to 3' direction, phosphoramidite synthesis proceeds in the 3' to 5' direction. The first monomer is attached via its 3' O to a solid support such as a glass bead, and its 5' O is initially protected from nonspecific reactions by conjugation of a dimethyloxytrityl (DMT) group.

**Figure 9: Phosphoramidite Reaction Cycle**



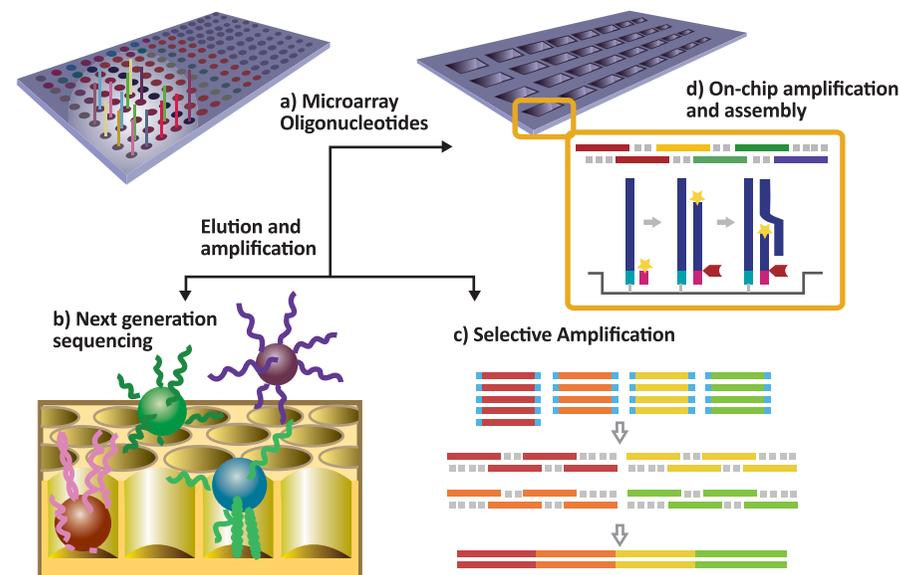
In **Step 1: Deprotection**, DMT is removed by washing with a mild acid such as trichloroacetic acid, exposing the 5' O for reaction. The second nucleotide is introduced to the reaction with its own 5' O protected with DMT, while its 3' O is activated by the conjugated phosphoramidite group. **Step 2: Coupling** occurs when the 3' O of the second nucleotide forms a phosphate triester bond with the 5' O of the first nucleotide. **Step 3: Capping** involves acetylating the 5' OH of any unreacted nucleotides to prevent later growth of an incorrect sequence. Acetic anhydride and dimethylaminopyridine are typically added to acetylate any unreacted terminal 5' OH groups, but will have no effect on terminal nucleotides that are protected by DMT. **Step 4: Oxidation** occurs upon the addition of iodine to convert the phosphate triester bond into the phosphodiester bond that forms the familiar backbone of DNA.

Once the desired chain is complete, all of the protecting groups must be removed, and the 5' end of the oligo is phosphorylated. Prematurely terminated strands can be removed by purifying the eluted product via gel electrophoresis and cutting out the band with the correct length. Further characterization of synthetic oligonucleotides may include sequencing or simply verifying the predicted molecular mass by mass spectrometry.

Oligo synthesis can be done in your lab in a column or plate format with the help of an automated synthesizer, such as a MerMade machine (BioAutomation). Alternatively, oligos may be cheaply and quickly synthesized by commercial vendors, including GenScript. Oligos serve not only as the building blocks to be assembled but also as primers to aid in assembly and PCR amplification of synthetic gene products.

New technologies are emerging to make high-throughput gene synthesis more rapid and affordable, through the use of microfluidics and microarray platforms (Figure 6).<sup>30</sup>

**Figure 10: High-throughput array-based gene synthesis technology**



Reprinted, with permission, from Ma, S., Tang, N. & Tian, J. DNA synthesis, assembly and applications in synthetic biology. *Current Opinion in Chemical Biology* **16**, 260–267 (2012).

Microarray technologies require some method of spatially confining reagents and/or directing reactions with very fine spatial and temporal control. Such methods include: using inkjet printing to dispense picoliter reagents to specific locations on a silica chip; controlling the deblocking step with light-activated photochemistry in a microfluidics system; or using programmable microelectrode arrays to direct redox reactions at desired spots. In practice, these methods can create error through imprecision and can introduce new costs, such as the expense of manufacturing unique photomasks required for photolithography platforms.

Microarray oligo pools provide a cheaper source of oligos, but make gene assembly more challenging and error-prone.<sup>30</sup> To overcome this challenge, highly parallel and miniaturized methods, such as “megacloning,” are being developed that combine the error-prone synthesis of oligo pools with a high-throughput pyrosequencing platform to enable the identification and selective amplification of oligos with the correct sequence prior to assembly.<sup>31</sup> Careful design of oligos for hybridization selection, or of primers for selective amplification of correct sequences, can allow direct assembly from unpurified oligo pools. These emerging technologies will eliminate the expensive and time-consuming oligo purification steps once the design algorithms and array-based platforms become widely available.

*GenPlus™ Next-Generation 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!*

### Step 3: Gene Assembly

Many methods of assembling oligos into complete genes or larger genome building blocks have been developed and successfully used (See table 5). For relatively short sequences (up to 1 kb), polymerase-based or ligase-based *in vitro* assembly methods are sufficient. For longer sequences, *in vivo* recombination-based methods may be preferred. Correct assembly requires a high-fidelity enzyme (e.g. DNA polymerase or ligase).

**Polymerase chain assembly (PCA)** is a standard technique for polymerase-based oligo assembly in a thermocycler. This reaction is also called templateless PCR. The principle is to combine all of the single-stranded oligos into a single tube, perform

thermocycling to facilitate repeated rounds of annealing, extension, denaturation, and then use the outermost primers to amplify full-length sequences. The success of this method depends upon the accurate synthesis of oligos designed to possess sufficient regions of overlap, sufficiently similar melting and annealing temperatures, and minimal opportunities for mis-hybridization. Protocols for PCA to produce final sequences up to ~750 kb have been published.<sup>32</sup>

**Table 5: Published Methods for Assembly of Oligos into Synthetic Genes or Genomes**

Oligo Assembly Methods	Reference
<b>POLYMERASE-BASED</b>	
Dual-Asymmetric (DA) PCR	Sandhu <i>et al.</i> (1992) <sup>33</sup>
Overlap Extension (OE)	Horton <i>et al.</i> (1993) <sup>34</sup>
Polymerase Cycling Assembly	Stemmer <i>et al.</i> (1995) <sup>35</sup>
Asymmetric PCR	Wooddell & Burgess (1996) <sup>36</sup>
Thermodynamically-Balanced Inside-Out (TBIO)	Gao <i>et al.</i> (2003) <sup>37</sup>
Two-Step (DA+OE)	Young & Dong (2004) <sup>38</sup>
Microchip-Based Multiplex Gene Synthesis	Tian <i>et al.</i> (2004) <sup>39</sup>
One-Step Simplified Gene Synthesis	Wu <i>et al.</i> (2006) <sup>40</sup>
Parallel Microfluidics-Based Synthesis	Kong <i>et al.</i> (2007) <sup>41</sup>
Single Molecule PCR	Yehezkel <i>et al.</i> (2008) <sup>42</sup>
TopDown Real-Time Gene Synthesis	Huang <i>et al.</i> (2012) <sup>43</sup>
<b>LIGASE-BASED</b>	
Shotgun Ligation	Eren & Swenson (1989) <sup>44</sup>
Two-Step Ligation and PCR	Mehta <i>et al.</i> (1997) <sup>45</sup>
Ligase Chain Reaction	Au <i>et al.</i> (1998) <sup>46</sup>
Brick-Based	Kelly <i>et al.</i> (2009) <sup>47</sup>
<b>RECOMBINATION-BASED</b>	
Sequence-and Ligation-Independent Cloning (SLIC)	Li and Elledge (2007) <sup>48</sup>
Transformation-Associated Recombination	Gibson <i>et al.</i> (2008) <sup>49</sup>
BioBrick Assembly in <i>E. coli</i>	Ho-shing <i>et al.</i> (2012) <sup>50</sup>

**Ligase Chain Reaction (LCR)** uses a DNA ligase to join overlapping ends of synthetic oligos in repeated cycles of denaturation, annealing, and ligation. Using a thermostable enzyme such as *Pfu* DNA ligase 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 inefficient as oligo number increases, limiting its usefulness to genes shorter than 2 kb.<sup>51</sup> Further, the use of standard-purity oligonucleotides can introduce point deletions resulting from incorporation of incorrectly synthesized oligos; this can be avoided by using gel-purified oligos, which increases costs substantially, or by performing side-directed mutagenesis to correct errors after assembly, cloning, and sequencing.<sup>55</sup>

**Sequence- and Ligation-Independent Cloning (SLIC)** is a method of *in vitro* homologous recombination employing a T4 DNA polymerase that allows the assembly of up to five gene fragments via simultaneous incorporation into a plasmid vector.<sup>47</sup> Synthetic oligos are prepared for SLIC by PCR extension to introduce flanking regions of sequence homology; 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. Homologous regions are annealed *in vitro* and undergo gap repair after transformation into *E. coli*.

***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 oligos. Gibson and colleagues have demonstrated that the yeast *Saccharomyces cerevisiae* can take up and assemble at least 38 overlapping single-stranded oligos 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 kilobase-sized synthetic DNA molecules. A protocol for designing the oligonucleotides to be assembled, transforming them into yeast, and confirming their assembly has been published.<sup>52</sup>

*GenScript's Gene-Brick™ synthesis service delivers 8-13 kb full-length, synthetic DNA fragments with 100% guaranteed sequence accuracy. Gene-Bricks are essential tools for synthetic biology projects such as in the Synthetic Yeast Genome (p. 17).*

## Step 4: Sequence Verification and Error Correction

Due to the inherent potential for error in each step of gene synthesis, all synthetic sequences should be verified before use. Sequences harboring mutations must be identified and removed from the pool or corrected. Internal insertions and deletions, as well as premature termination, are common in synthetic DNA sequences. The accumulation of errors from phosphoramidite chemical synthesis alone can lead to only about 30% of any synthesized 100-mer being the desired sequence.<sup>53</sup> Improper annealing during oligo assembly can also introduce heterogeneity in the final pool of synthetic gene products.

Cloning newly synthesized sequences into a plasmid vector can simplify the process of sequence verification. 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.

In the event that the correct sequence cannot be obtained and amplified from the pool of synthesized DNA, numerous methods for error detection and correction have been developed and successfully used. These include stringent hybridization using carefully designed oligos; exhaustive purification using electrophoresis, mass spectrometry and other biochemical methods; mismatch-binding or mismatch-cleavage using prokaryotic endonucleases; selection of correct coding sequences via functional assays; and site-directed mutagenesis after sequencing.<sup>53, 54</sup> PAGE or agarose gel purification is the technique in longest use, but it is costly and labor intensive and does not identify or correct for substitutions or for small insertions/deletions. Enzyme-based strategies are limited by the enzyme's capabilities; for example, the widely used *E. coli* MutHLS is not very effective for substitutions other than G-T, A-C, G-G, A-A.<sup>54</sup> Site-directed mutagenesis after sequencing can introduce point mutations using mutant primers and high-fidelity DNA polymerase followed by selection for unmethylated molecules, but can be an unwieldy technique. Because error correction can be so time-consuming and costly, especially for long or complex sequences, efforts continue to improve the accuracy of oligo synthesis and assembly.

*GenScript's Gene-on-Demand® gene synthesis platform enables mutation-free DNA synthesis for genes of any length and any complexity. Sequence chromatograms are delivered with all gene synthesis orders, and 100% sequence accuracy is always guaranteed!*

## Step 5: Preparing Synthetic DNA for Downstream Applications

### Cloning

For most applications, synthetic genes need to be cloned into appropriate vectors. These may include plasmid vectors for transfection or electroporation into cells, or viral vectors (such as adenovirus, retrovirus, or lentivirus) for transduction into cells or live animals.

To facilitate cloning, synthetic genes may be designed to include restriction enzyme sites, recombination arms, or other flanking sequences. As an alternative to avoid introducing special flanking sequences, recombination-based methods begin with PCR extension of the gene insert to introduce 15 bases of sequence homology to the linearized vector, facilitating homologous recombination without adding unwanted bases.

*GenScript's CloneEZ™ recombination method uses a high-fidelity enzyme to produce directional cloning of any gene into any position in any vector in only 30 minutes.*

### Propagation of Synthetic Genes

Propagation of synthetic gene constructs may be challenging for several reasons. Some “low copy number” plasmid constructs simply don't amplify well in commonly used bacterial propagation hosts. Long genes are often difficult 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. Genes may be toxic or unstable in certain host cells but not others. Screening a number of cell lines to find the best host for propagation can be time consuming, but may be necessary for certain difficult sequences.

*GenScript's Elite System™ Host Cell Pool can determine the best bacterial host for propagating difficult genes and deliver a vial of these cells along with the purified plasmid DNA.*

### Plasmid Preparation Quality

Depending on your research needs, the quality of plasmid obtained from gene synthesis project may need varying levels of quality control. While there are many factors that need to be considered when evaluating plasmid prep quality, such as A260/280 and residual RNA, two of the main factors are endotoxin levels and supercoiling percentages

Supercoiled plasmids have a better transformation efficiency than plasmids that are relaxed. Relaxed plasmids are transformed with 25% less efficiency than supercoiled plasmids. Thus, higher levels of supercoiling in a plasmid prep results in a higher subsequent transformation efficiency.

The other factor that needs to be considered is endotoxin levels. Endotoxins are molecules that are comprised of a lipid and a polysaccharide. They 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 in some manufacturing products. 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 manufacturing
- Animal studies

*GenScript's Transfection Grade Plasmid Preps offer industry-leading quality control, with the lowest endotoxin levels and highest supercoiling percentages for guaranteed success in downstream applications such as transfection of mammalian cells.*

### Biosafety

Throughout the history of genetics research, numerous regulatory boards have addressed the concerns of researchers, citizens, and political leaders regarding the potential for genetic technologies to be used for harm. In 1974, the NIH formed a

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.



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. This group works 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.

## The Future of Gene Synthesis

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 \$0.35 per bp over the last decade, in parallel with advances in DNA sequencing and chip-based bioassays, and consistent with Moore's Law. Further improvements in automation, error correction, and cost-effectiveness will allow gene synthesis to become a valuable 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 for regulating biosynthesis 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 discussion by the research community, biotech industry, and international governmental and regulatory bodies. Without a doubt, the coming decade will present exciting opportunities for gene synthesis to continue to fuel innovation in the life sciences.

## References

1. Michelson, A. M. & Todd, A. R. Nucleotides part XXXII. Synthesis of a dithymidine dinucleotide containing a 3': 5'-internucleotidic linkage. *Journal of the Chemical Society (Resumed)* 2632–2638 (1955).
2. Agarwal, K. L. *et al.* Total synthesis of the gene for an alanine transfer ribonucleic acid from yeast. *Nature* **227**, 27–34 (1970).
3. Rothstein, R. J. One-step gene disruption in yeast. *Methods in enzymology* **101**, 202–211 (1983).
4. Kimmel A. R. & Faix J. Generation of multiple knockout mutants using the Cre-loxP system. *Methods Mol Biol.* 346:187-99 (2006).
5. Gaj, T., Gersbach, C. A. & Barbas, C. F., 3rd. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol.* **31**, 397–405 (2013).
6. Wang, L. & Mu, F. Y. A Web-based design center for vector-based siRNA and siRNA cassette. *Bioinformatics* **20**, 1818–1820 (2004).
7. Yousefi, S. *et al.* Catapult-like release of mitochondrial DNA by eosinophils contributes to antibacterial defense. *Nat. Med.* **14**, 949–953 (2008).
8. Fuchs, T. A. *et al.* Extracellular DNA traps promote thrombosis. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 15880–15885 (2010).
9. Wen, F., Shen, A., Choi, A., Gerner, E. W. & Shi, J. Extracellular DNA in pancreatic cancer promotes cell invasion and metastasis. *Cancer Res.* **73**, 4256–4266 (2013).
10. Gloag, E. S. *et al.* Self-organization of bacterial biofilms is facilitated by extracellular DNA. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 11541–11546 (2013).
11. Cantini *et al.* Identification and characterization of Dicer1e, a Dicer1 protein variant, in oral cancer cells. *Mol. Cancer* **13**: 190 (2014).
12. Stewart-Jones *et al.* Trimeric HIV-1-Env structures define glycan shields from clades A, B, and G. *Cell* **165**: 1-14 (2016).
13. Joyce *et al.* Vaccine-induced antibodies that neutralize Group 1 and Group 2 Influenza A Viruses. *Cell* **166**: 609-623 (2016).
14. Hutchison *et al.* Design and synthesis of a minimal bacterial genome. *Science* **351**: 6280 (2016).
15. Tianpei *et al.* Scorpion peptide LqhIT2 activates phenylpropanoid pathways via jasmonate to increase rice resistance to rice leafrollers. *Plant Sci.* **230**: 1-11 (2015).
16. Burgess-Brown, N. A. *et al.* Codon optimization can improve expression of human genes in *Escherichia coli*: A multi-gene study. *Protein Expr. Purif.* **59**, 94–102 (2008).
17. Yoo, S. K., Starnes, T. W., Deng, Q. & Huttenlocher, A. Lyn is a redox sensor that mediates leukocyte wound attraction *in vivo*. *Nature* **480**, 109–112 (2011).
18. Xiong, A.-S. *et al.* A semi-rational design strategy of directed evolution combined with chemical synthesis of DNA sequences. *Biol. Chem.* **388**, 1291–1300 (2007).
19. Dimitrov, D. S. Therapeutic proteins. *Methods Mol. Biol.* **899**, 1–26 (2012).
20. Brazier-Hicks, M. & Edwards, R. Metabolic engineering of the flavone-C-glycoside pathway using polyprotein technology. *Metab. Eng.* **16**, 11–20 (2013).
21. Dellomonaco, C., Rivera, C., Campbell, P. & Gonzalez, R. Engineered respiro-fermentative metabolism for the production of biofuels and biochemicals from fatty acid-rich feedstocks. *Appl. Environ. Microbiol.* **76**, 5067–5078 (2010).
22. Friedland, A. E. *et al.* Heritable genome editing in *C. elegans* via a CRISPR-Cas9 system. *Nature Methods* **10**, 741–743 (2013).
23. Dymond, J. & Boeke, J. The *Saccharomyces cerevisiae* SCRaMbLE system and genome minimization. *Bioengineered Bugs* **3**, 5–4 (2012).
24. Dymond, J. S. *et al.* Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. *Nature* **477**, 471–476 (2011).
25. Hoover, D. M. & Lubkowsky, J. DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. *Nucleic Acids Res.* **30**, e43 (2002).
26. Hoover, D. Using DNAWorks in designing oligonucleotides for PCR-based gene synthesis. *Methods Mol. Biol.* **852**, 215–223 (2012).
27. Rouillard, J.-M. *et al.* Gene2Oligo: oligonucleotide design for *in vitro* gene synthesis. *Nucleic Acids Res.* **32**, W176–180 (2004).

28. Richardson, S. M., Wheelan, S. J., Yarrington, R. M. & Boeke, J. D. GeneDesign: rapid, automated design of multikilobase synthetic genes. *Genome Res.* **16**, 550–556 (2006).
29. Bode, M., Khor, S., Ye, H., Li, M.-H. & Ying, J. Y. TmPrime: fast, flexible oligonucleotide design software for gene synthesis. *Nucleic Acids Res.* **37**, W214–221 (2009).
30. Ma, S., Tang, N. & Tian, J. DNA synthesis, assembly and applications in synthetic biology. *Current Opinion in Chemical Biology* **16**, 260–267 (2012).
31. Matzas, M. *et al.* High-fidelity gene synthesis by retrieval of sequence-verified DNA identified using high-throughput pyrosequencing. *Nature Biotechnology* **28**, 1291–1294 (2010).
32. Marchand, J. A. & Peccoud, J. Building block synthesis using the polymerase chain assembly method. *Methods Mol. Biol.* **852**, 3–10 (2012).
33. Sandhu, G. S., Aleff, R. A. & Kline, B. C. Dual asymmetric PCR: one-step construction of synthetic genes. *BioTechniques* **12**, 14–16 (1992).
34. Horton, R. M. *et al.* Gene splicing by overlap extension. *Meth. Enzymol.* **217**, 270–279 (1993).
35. Stemmer, W. P., Cramer, A., Ha, K. D., Brennan, T. M. & Heyneker, H. L. Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene* **164**, 49–53 (1995).
36. Wooddell, C. I. & 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.* **6**, 886–892 (1996).
37. Gao, X., Yo, P., Keith, A., Ragan, T. J. & Harris, T. K. 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.* **31**, e143 (2003).
38. Young, L. & Dong, Q. Two-step total gene synthesis method. *Nucleic Acids Res.* **32**, e59 (2004).
39. Tian, J. *et al.* Accurate multiplex gene synthesis from programmable DNA microchips. *Nature* **432**, 1050–1054 (2004).
40. Wu, G. *et al.* Simplified gene synthesis: a one-step approach to PCR-based gene construction. *J. Biotechnol.* **124**, 496–503 (2006).
41. Kong, D. S., Carr, P. A., Chen, L., Zhang, S. & Jacobson, J. M. Parallel gene synthesis in a microfluidic device. *Nucleic Acids Res.* **35**, e61 (2007).
42. Ben Yehezkel, T. *et al.* De novo DNA synthesis using single molecule PCR. *Nucleic Acids Res.* **36**, e107 (2008).
43. Huang, M. C., Cheong, W. C., Ye, H. & Li, M.-H. TopDown real-time gene synthesis. *Methods Mol. Biol.* **852**, 23–34 (2012).
44. Eren, M. & Swenson, R. P. Chemical synthesis and expression of a synthetic gene for the flavodoxin from *Clostridium MP*. *J. Biol. Chem.* **264**, 14874–14879 (1989).
45. Mehta, D. V., DiGate, R. J., Banville, D. L. & Guiles, R. D. Optimized gene synthesis, high level expression, isotopic enrichment, and refolding of human interleukin-5. *Protein Expr. Purif.* **11**, 86–94 (1997).
46. Au, L. C., Yang, F. Y., Yang, W. J., Lo, S. H. & Kao, C. F. Gene synthesis by a LCR-based approach: high-level production of leptin-L54 using synthetic gene in *Escherichia coli*. *Biochem. Biophys. Res. Commun.* **248**, 200–203 (1998).
47. Kelly, J. R. *et al.* Measuring the activity of BioBrick promoters using an *in vivo* reference standard. *J Biol Eng* **3**, 4 (2009).
48. Li, M. Z. & Elledge, S. J. SLIC: a method for sequence- and ligation-independent cloning. *Methods Mol. Biol.* **852**, 51–59 (2012).
49. Gibson, D. G. *et al.* Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science* **319**, 1215–1220 (2008).
50. Ho-Shing, O., Lau, K. H., Vernon, W., Eckdahl, T. T. & Campbell, A. M. Assembly of standardized DNA parts using BioBrick ends in *E. coli*. *Methods Mol. Biol.* **852**, 61–76 (2012).
51. Chalmers, F. M. & Curnow, K. M. Scaling up the ligase chain reaction-based approach to gene synthesis. *BioTechniques* **30**, 249–252 (2001).
52. Gibson, D. G. Oligonucleotide assembly in yeast to produce synthetic DNA fragments. *Methods Mol. Biol.* **852**, 11–21 (2012).
53. Hughes, R. A., Miklos, A. E. & Ellington, A. D. Gene synthesis: methods and applications. *Meth. Enzymol.* **498**, 277–309 (2011).
54. Ma, S., Saaem, I. & Tian, J. Error correction in gene synthesis technology. *Trends Biotechnol.* **30**, 147–154 (2012).

## New Service Features!

### Now part of our comprehensive MB portfolio

#### HT Plasmid Prep Service

- Latest addition to our plasmid prep line allows for miniprep quantities of prep (10 µg and 30 µg offerings)
- No minimum on the number of preps required
- Still maintain high-quality QC parameters seen with our standard plasmid prep service
- HT Plasmid Prep Applications:
  - Multi-well plate-based assays, ELISA
  - Antibody expression and affinity assays
  - Structural Biology Studies
  - High-throughput sequencing
  - Probe generation

#### Coming Soon! Express Plasmid Prep Service

- Exclusive service line for plasmid prep that is available for bundling with gene synthesis
- Take advantage of our industry-leading gene synthesis service while decreasing plasmid prep turnaround times to as little as 2 business days
- Available in research grade and transfection grade QC levels.

# GenScript's Molecular Biology Service Portfolio

## Gene Synthesis

### Gene Synthesis

*Any DNA sequence you desire*

- Guaranteed 100% sequence fidelity
- Free OptimumGene™ codon optimization

### Express Gene Synthesis

*Speed comes from experience*

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

### GenPlus™

*Next-generation gene synthesis*

- Lowest price for high-throughput gene synthesis (25 genes or more)
- Unprecedented capacity of 100 million bp/month

### GenBrick™ Synthesis

*Building blocks for synthetic biology*

- 8-15 kb long DNA constructs
- \$0.45/bp, 23 business days

## Mutagenesis & Variant Libraries

### Express Mutagenesis

*Any variant from your design*

- Starting from \$99/mutation
- 5 business days
- 100% sequence accuracy guaranteed

### Mutant Libraries

*You design it, we build it*

- Site-directed mutagenesis libraries
- Degenerated libraries
- Randomized libraries
- Customized libraries

### Gene Variant Libraries

*Up to 2,000 unique DNA sequences in 30 days*

- Combinatorial assembly libraries for synthetic biology
- Truncation variant libraries for structural biology
- Customized libraries for any research application

### High-Throughput Gene to Antibody

*Accelerate your antibody drug screening project*

- Industry-leading capacity
- Cost efficient: starting from \$399/rAb
- Up to 1mg recombinant antibody as fast as 3 weeks

**NEW!**

## Express Services Portfolio – Gene Synthesis Bundle Options

*Accelerate your research with one-stop downstream solutions*

### 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

### Coming Soon! 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 2 business days when bundled with gene synthesis

## 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

### Plasmid Preparation

*Any amount to fit your needs*

- Flexible scale up to gram level
- ≥ 95% Supercoiled DNA
- ≤ 0.005 EU/μg Endotoxin

### DNA Sequencing (North America)

*Rapid sequencing at low prices*

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

### Express Cloning

*Save time with clone-ready construct*

- \$49 and 2 days turnaround
- Choose from over 150 free vectors

### Custom Oligos

*Affordable modified/labeled oligos*

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