Artificial gene synthesis

Artificial gene synthesis or gene synthesis, refers to a group of methods that are used in synthetic biology to construct and assemble genes from nucleotides *de novo*. Unlike DNA synthesis in living cells, artificial gene synthesis does not require template DNA, allowing virtually any DNA sequence to be synthesized in the laboratory.
It comprises two main steps,

- The first of which is solidphase DNA synthesis, sometimes known as 'DNA printing'.This produces oligonucleotide fragments that are generally under 200 base pairs.
- The second step then involves connecting these oligonucleotide fragments using various DNA assembly methods.

Because artificial gene synthesis does not require template DNA, it is theoretically possible to make a completely synthetic DNA molecules with no limits on the nucleotide sequence or size. Synthesis of the first complete gene, a yeast tRNA, was demonstrated by Har Gobind Khorana and coworkers in 1972. Synthesis of the first peptide- and protein-coding genes was performed in the laboratories of Herbert Boyer and Alexander Markham, respectively. More recently, artificial gene synthesis methods have been developed that will allow the assembly of entire chromosomes and genomes. The first synthetic yeast chromosome was synthesised in 2014, and entire functional bacterial chromosomes have also been synthesised. In addition, artificial gene synthesis could in the future make use of novel nucleobase pairs (unnatural base pairs).

Standard methods for DNA synthesis
Oligonucleotide synthesis
Annealing based connection of oligonucleotides
Limitations
Error correction procedures
Unnatural base pairs

Oligonucleotide synthesis

Oligonucleotides are chemically synthesized using building blocks called nucleoside phosphoramidites. These can be normal or modified nucleosides which have protecting groups to prevent their amines, hydroxyl groups and phosphate groups from interacting incorrectly. One phosphoramidite is added at a time, the 5' hydroxyl group is deprotected and a new base is added and so on. The chain grows in the 3' to 5'direction, which is backwards relative to biosynthesis. At the end, all the protecting groups are removed. Nevertheless, being a chemical process, several incorrect interactions occur leading to some defective products. The longer the oligonucleotide sequence that is being synthesized, the more defects there are, thus this process is only practical for producing short sequences of nucleotides. The

current practical limit is about 200bp (base pairs) for an oligonucleotide with sufficient quality to be used directly for a biological application.

Usually, a set of individually designed oligonucleotides is made on automated solid-phase synthesizers, purified and then connected by specific annealing and standard ligation or polymerase reactions. To improve specificity of oligonucleotide annealing, the synthesis step relies on a set of thermostable DNA ligase and polymerase enzymes.

Now a days, several methods for gene synthesis have been described, such as the

- ligation of phosphorylated overlapping oligonucleotides,
- the Fok I method and
- A modified form of ligase chain reaction for gene synthesis.

Additionally, several PCR assembly approaches have been described. They usually employ oligonucleotides of 40-50 nucleotides long that overlap each other. These oligonucleotides are designed to cover most of the sequence of both strands, and the full-length molecule is generated progressively by overlap extension (OE) PCR, thermodynamically balanced inside-out (TBIO) PCR or combined approaches.

Moreover, because the assembly of the full-length gene product relies on the efficient and specific alignment of long single stranded oligonucleotides, critical parameters for synthesis success include extended sequence regions comprising secondary structures caused by inverted repeats, extraordinary high or low GC-content,or repetitive structures. Usually these segments of a particular gene can only be synthesized by splitting the procedure into several consecutive steps and a final assembly of shorter sub-sequences, which in turn leads to a significant increase in time and labor needed for its production. The result of a gene synthesis experiment depends strongly on the quality of the oligonucleotides used. For these annealing based gene synthesis protocols, the quality of the product is directly and exponentially dependent on the correctness of the employed oligonucleotides.

Error correction procedures

Oligonucleotide quality several elaborate strategies have been developed, employing either separately prepared fishing oligonucleotides, mismatch binding enzymes of the mutS family or specific endonucleases from bacteria or phages. Nevertheless, all these strategies increase time and costs for gene synthesis based on the annealing of chemically synthesized oligonucleotides.

Increasingly, genes are ordered in sets including functionally related genes or multiple sequence variants on a single gene. Virtually all of the therapeutic proteins in development, such as monoclonal antibodies, are optimised by testing many gene variants for improved function or expression, before massively parallel sequencing. Tag-directed primers then enable the retrieval of molecules with desired sequences by dial-out PCR.

Increasingly, genes are ordered in sets including functionally related genes or multiple sequence variants on a single gene. Virtually all of the therapeutic proteins in development, such as monoclonal antibodies, are optimised by testing many gene variants for improved function or expression. While traditional nucleic acid synthesis only uses 4 base pairs -

adenine, thymine, guanine and cytosine, oligonucleotide synthesis in the future could incorporate the use of unnatural base pairs, which are artificially designed and synthesized nucleobases that do not occur in nature.

Error correction procedures

Unnatural base pairs

algal gene that expresses a nucleotide triphosphate transporter which efficiently imports the triphosphates of both d5SICSTP and dNaMTP into *E. coli* bacteria. Then, the natural bacterial replication pathways use them to accurately replicate the plasmid containing d5SICS–dNaM. The successful incorporation of a third base pair is a significant breakthrough toward the goal of greatly expanding the number of amino acids which can be encoded by DNA, from the existing 20 amino acids to a theoretically possible 172, thereby expanding the potential for living organisms to produce novel proteins.[20] In the future, these unnatural base pairs could be synthesised and incorporated into oligonucleotides via DNA printing methods.

Polymerase Chain Reaction (PCR) :

Principle, Procedure, Components, Types , Applications and PCR Stages

The polymerase chain reaction (PCR) is a laboratory technique for DNA replication that allows a "target" DNA sequence to be selectively amplified. PCR can use the smallest sample of the DNA to be cloned and amplify it to millions of copies in just a few hours. Discovered in 1985 by Kerry Mullis, PCR has become both and essential and routine tool in most biological laboratories.

Principle of PCR

The PCR involves the **prime r mediated enzymatic amplification** of DNA. PCR is based on using the ability of DNA polymerase to synthesize new strand of DNA complementary to the offered template strand. Primer is needed because DNA polymerase can add a nucleotide only onto a preexisting 3′-OH group to add the first nucleotide. DNA polymerase then elongate its 3 end by adding more nucleotides to generate an extended region of double stranded DNA.

Components of PCR

The PCR reaction requires the following components:

1. DNA Template : The double stranded DNA (dsDNA) of interest, separated from the sample.

2. DNA Polymerase : Usually a thermostable Taq polymerase that does not rapidly denature at high temperatures (98°), and can function at a temperature optimum of about 70°C.

3. Oligonucleotide prime rs : Short pieces of single stranded DNA (often 20-30 base pairs) which are complementary to the 3' ends of the sense and anti-sense strands of the target sequence.

4. Deoxynucleotide triphosphates : Single units of the bases A, T, G, and C (dATP, dTTP, dGTP, dCTP) provide the energy for polymerization and the building blocks for DNA synthesis.

5. Buffer system : Includes magnesium and potassium to provide the optimal conditions for

DNA denaturation and renaturation; also important for polymerase activity, stability and fidelity.

Procedure of PCR

All the PCR components are mixed together and are taken through series of 3 major cyclic reactions conducted in an automated, self-contained thermocycler machine.

1. Initialization Step:

It is the first step of the cycle which consists of raising the temperature of the reaction to 94–96°C or 98 °C if extremely thermostable polymerases are used, which is held for 1–9 minutes. This process activates the DNA polymerase used in the reaction.

2. Denaturation :

This step involves heating the reaction mixture to 94°C for 15-30 seconds. During this, the double stranded DNA is denatured to single strands due to breakage in weak hydrogen bonds.

3. Annealing :

The reaction temperature is rapidly lowered to 54-60°C for 20-40 seconds. This allows the primers to bind (anneal) to their complementary sequence in the template DNA.

4. Elongation :

Also known at extension, this step usually occurs at 72-80°C (most commonly 72°C). In this step, the polymerase enzyme sequentially adds bases to the 3′ each primer, extending the DNA sequence in the 5′ to 3′ direction. Under optimal conditions, DNA polymerase will add about 1,000 bp/minute.

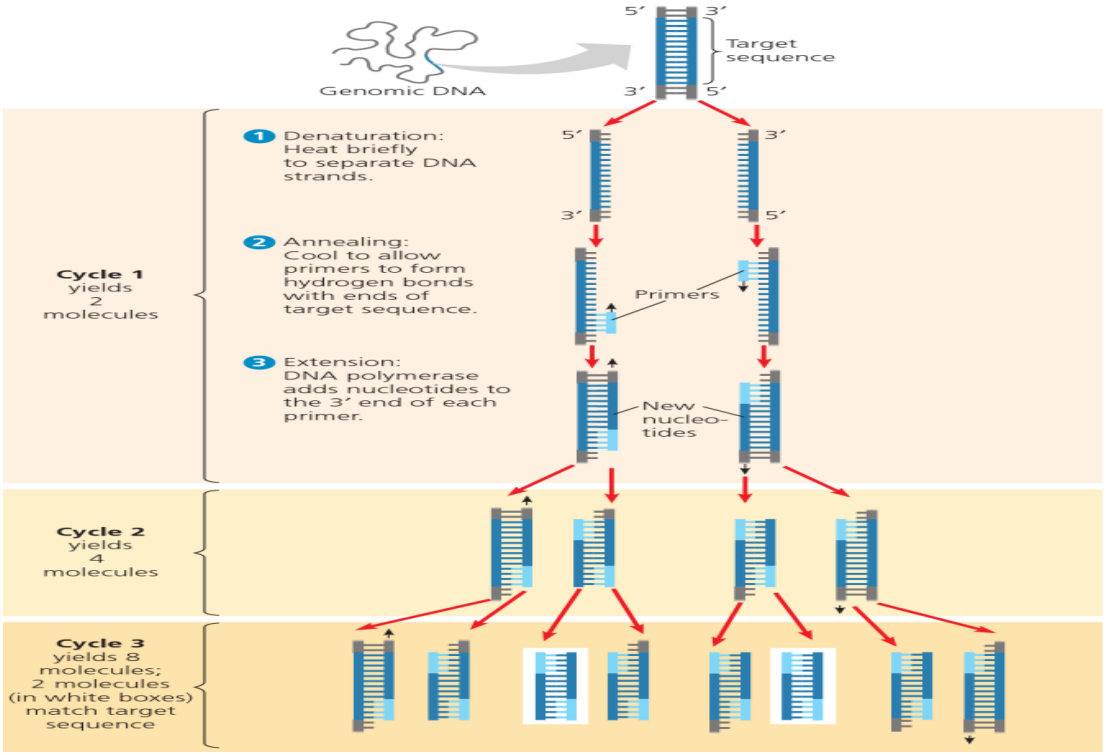
With one cycle, a single segment of double-stranded DNA template is amplified into two separate pieces of double-stranded DNA. These two pieces are then available for amplification in the next cycle. As the cycles are repeated, more and more copies are generated and the number of copies of the template is increased exponentially.

5. Final elongation.

This step is performed at a temperature of 70-74 degree centigrade for 5-15 minutes after the last PCR cycle to ensure that any remaining single-stranded DNA is fully extended.

6. Final hold.

In this step the mixture is allowed to cool to a temperature of 4-15 degree centigrade for short term storage of the reaction.



Types of PCR

In addition to the amplification of a target DNA sequence by the typical PCR procedures already described, several specialised types of PCR have been developed for specific applications.

1. Real-time PCR
2. Quantitative real time PCR (Q-RT PCR)
3. Reverse Transcriptase PCR (RT-PCR)
4. Multiplex PCR
5. Nested PCR
6. Long-range PCR
7. Single-cell PCR
8. Fast-cycling PCR
9. Methylation-specific PCR (MSP)
10. Hot start PCR
11. High-fidelity PCR
12. In situ PCR
13. Variable Number of Tandem Repeats (VNTR) PCR
14. Asymmetric PCR
15. Repetitive sequence-based PCR
16. Overlap extension PCR
17. Assemble PCR
18. Intersequence-specific PCR(ISSR)
19. Ligation-mediated PCR
20. Methylation –specifin PCR
21. Miniprimer PCR

22. Solid phase PCR
23. Touch down PCR, etc

Applications of PCR
Some common applications of PCR in various fields can be explained in following categories.

Medical Applications:

1. Genetic testing for presence of genetic disease mutations. Eg: hemoglobinopathies, cysticfibrosis, other inborn errors of metabolism.

2. Detection of disease causing genes in suspected parents who act as carriers. Study of alteration to oncogenes may help in customization of therapy

3. Can also be used as part of a sensitive test for tissue typing, vital to organ transpla ntation genotyping of embryo **4.** Helps to monitor the gene in gene therapy

Infectious disease Applications:

- Analyzing clinical specimens for the presence of infectious agents, including HIV, hepatitis, malaria, tuberulosis etc.
- Detection of new virulent subtypes of organism that is responsible for epidemics.

Forensic Applications:

Can be used as a tool in genetic fingerprinting. This technology can identify any one person from millions of others in case of : crime scence, rule out suspects during police investigation,paternity testing even in case of avaibility of very small amount of specimens (stains of blood,semen, hair etc)

Research and Molecular Genetics:

1. In genomic studies: PCR helps to compare the genomes of two organisms and identify the difference between them.
2. In phylogenetic analysis. Minute quantities of DNA from any source such a fossilized material, hair, bones, mummified tissues.
3. In study of gene expression analysis, PCR based mutagenesis
4. In Human genome project for aim to complete mapping and understanding of all genes of human beings.

Components

The basic components and reagents required to set up a 100 ul PCR reaction are:

1. Microfuge tube.

These are small cylindrical plastic conical containers with conical bottoms with a snap cap. They are made up of polypropylene, thus can withstand a wide range of temperature.

2. Thermal cycler.

3. It is an apparatus used to amplify segments of DNA. It has a thermal block with holes where tubes holding the PCR reaction mixtures can be inserted. The cycler works on the principle of Peltier effect, which raises and lowers the temperature of the block in a preprogrammed manner by reversing the electric current. Thin-walled reaction tubes permit favorable thermal conductivity to allow for rapid thermal equilibration.

4. DNA template.

The reaction solution should contain at least (1e5-1e6 target molecules).

5. Primer.

These are oligonucleotides that define the sequence to be amplified. Two primer that are complementary to the 3′ (three prime) ends of each of the sense and anti-sense strand of the DNA target (Tm 52-58 degree centigrade preferred). Primers with melting temperatures above 65 degree centigrade have a tendency for secondary annealing. The GC content (the number of G's and C's in the primer as a percentage of the total bases) of primer should be 40-60%.

Formula for primer Tm calculation:

Melting Temperature Tm(oK) = {ΔH/ ΔS + R ln(C)}, Or Melting Temperature Tm(oC) = {ΔH/ΔS + R ln(C)}-273.15 where

ΔH (kcal/mole): H is the Enthalpy. Enthalpy is the amount of heat energy possessed by substances. ΔH is the change in Enthalpy. I n the above formula the ΔH is obtained by adding up all the di-nucleotide pair enthalpy values of each nearest neighbor base pair.

ΔS (kcal/mole): S is the amount of disorder a system exhibits is called entropy. ΔS is change in Entropy. Here it is obtained by adding up all the di-nucleotide pair's entropy values of each nearest neighbor base pair. An additional salt correction is added as the Nearest Neighbor parameters were obtained from DNA melting studies conducted in 1M Na+ buffer and this is the default condition used for all calculations.

ΔS (salt correction) = ΔS (1M NaCl) + 0.368 x N x ln([Na+])

Where

N is the number of nucleotide pairs in the primer (primer length -1).

[Na+] is salt equivalent in mM.

The primer annealing temperature is defined by the formula:

Ta = 0.3 x Tm (primer) + 0.7 Tm (product) -14.9

where, Tm(primer) = Melting Temperature of the primers

Tm(product) = Melting temperature of the product

6. Tris-HCl.

The recommended buffer solution is 10 to 50 mM Tris-HCl (pH 8.3-8.8) at 20 degree centigrade.

7. *MgCl2*.

It is the cofactor of the enzyme. It is beneficial to optimize the magnesium ion concentration. The magnesium ion affects the primer annealing, strand dissociation temperatures of template and PCR product, product specificity, formation of primer-dimer artifacts and enzymatic activity and fidelity. Taq DNA polymerase requires free magnesium that binds to template DNA,primers, and dNTPs.

8. KCl.

KCl is to be used for the reaction to facilitate primer annealing.

9. Gelatin or bovine serum.

Aautoclaved gelatin or nuclease- free bovine serum albumins are included to help stabilize the enzyme.

10. Distilled water.

Autoclaved distilled water was used. The volume depends on the reaction.

11. Deoxyneuclotide triphosphates.

These are the DNA building blocks. Dntp (TTP-thymidine triphosphate), dCTP (deoxycyctidine triphosphate), dATP (deoxyadenosine triphosphate) and dGTP (deoxyguanosine triphosphate) solutions neutralized to pH 7.0. Primary stock solution are diluted to 10 mM, aliquoted, and stored at -20 degree C. A working stock containing 1 mM each dNTP is recommended. The stability of dNTPs during repeated cycles of PCR is such that approximately 50% remains as dNTP after 50 cycles (Corey Levevenson, personal communication). dNTP concentrations between 20 and 200 uM is best for the reaction. The 4 dNTP should be at equivalent concentrations to minimize mis-incorporation error.

12. DNA polymerase.

It is an enzyme that catalyzes the reaction. Taq DNA polymerase isolated from *Thermus aquaticus* growing in hot springs acts best at 72 degree centigrade and the denaturation temperature of 90 degree centigrade does not destroy its enzymatic activity. Other thermostable enzyme like Pflu DNA polymerase isolated from *Pyrococcus furiosus* and Vent polymerase isolated from *Thermococcus litoralis*, were discovered and were found to be more efficient. A recommended concentration of Taq polymerase (Perkin-Elmer Cetus) is between 1 and 2.5 units (SA=20 units/pmol) per 100 uL reaction. However enzymatic activity will vary with respect to individual target templates or primers.

To set up a 25 or 50 uL reaction the concentration of the reagent are as follows:

Reagent		
10X buffer	2.5 µl	5 µl
dNTP	0.5	1
Forward primer	1	2
Reverse primer	1	2
Taq polymerase	0.15	0.3
Water	18.85 µl	38.7 µl
DNA (30-50 ng)	1	1
Total volume	25 µl	50 µl

PCR Stages

1. Exponential amplification.

As a result of each cycle, the number of copies of the desired segment becomes twice the number present at the end of the previous cycle.

The more times the three PCR cycles are repeated the more DNA you can obtain. This is because every cycle of a PCR reaction theoretically doubles the amount of target copies, so we expect a geometric amplification. In other words PCR is an exponential process.

One could use this formula to calculate the theoretical output of any input:

Y = X (1 + efficiency) n Y=amount of amplification target

X=input copy number

n =number of cycles

Efficiency factor is given for each cycle in the kit

2. Leveling off stage.

The reaction slows as the DNA polymerase loses activity and as consumption of reagents such as dNTPs and primers causes them to become limiting.

3. Plateau stage.

The term "plateau effect" is used to describe the attenuation in the exponential rate of product accumulation that occurs during late PCR cycles. The plateau effect is affected by:

- Utilization of substrates (dNTPs or primers).
- Stability of reactants (dNTPs or enzyme).
- End-product inhibition (pyrophosphate, duplex DNA).
- Competition for reactants by nonspecific products or primer-dimer.
- Reannealing of specific product at concentrations above 1E8 M (may decrease the extension rate or processivity of Taq DNA polymerase or cause branch-migration of products strands and displacement of primers.Incomplete denaturation/strand separation of product at high concentration.