NovoPro FAQs

Q1: What options does NovoPro offer for gene deliverables?

A: Our standard deliverable is 5 µg lyophilized plasmid containing the gene of interest. An extra cost will apply if a bacstab or glycerol stock is requested. It is not recommended to directly use the 5 µg plasmid forcell transfection or mammalian protein expression. However, we can work with you on an additional plasmid to suit your project.

Q2: How to handle the lyophilized plasmid?

a. Centrifuge at 5,000×g for 5 min.
b. Carefully open the tube and add 20 µl of sterile water to dissolve the DNA.
c. Close the tube and incubate for 10 minutes at room temperature.
d. Briefly vortex the tube and then do a quick spin to concentrate the liquid at the bottom. Speed is less than 5000×g.
e. Store the plasmid at -20 ℃.

Q3: How to store gene deliverables, including lyophilized plasmid, bacstab and glycerol stock?

A: Lyophilized plasmid is stable under -20°C for at least one year, bac stab stable at 4°C for one week and glycerol stock at -80°C for several years.

For some of our gene library deliverables, liquid plasmid and/or PCR products need to be stored at -20°C and avoid repeated freezing and thawing.

Q4: What kind of bacterial strain is used in gene synthesis service, and how should it be handled?

A: Generally, we use *E.coli* TOP10 strain for most of our gene synthesis services. Whilst for special cases, we may apply Stbl3, EPI400 or EPI300 strains. The information of what competent cell type used in the service can be shared with our customer along with the corresponding bacstab or glycerol stock. Growth conditions to handle different cell type:

TOP10, growth in LB overnight at 37°C

Stbl3, growth in LB for 20~24 h at 30°C

EPI400/EPI300, the protocol is as below:

a) Add 4 ml LB media into each test tube. Inoculate each tube with bacterial culture with antibiotic at the proper concentration.
b) Incubate the tubes at 37°C, shaking overnight.
c) Dilute the starting culture (from step b) 1:10 into antibiotic supplemented fresh media.
d) Supplement induction solution with a ratio of 1:1000, grow the culture at 37°C for 4 h with vigorous shaking (approx. 250 rpm).
e) Isolate DNA from the induced culture cells as per protocol provided.

Q5: How should I amplify plasmid DNA from a bacstab or glycerol stock?

A: For bacstab or glycerol stock, we suggest the following steps:
a) streak bacteria on the plate
b) isolate a single colony
c) inoculate it in LB containing suitable antibiotic
d) plasmid isolation as usual

Q6: Some helpful tips on the plasmid handling?

A: In order to avoid any loss from the cap of the vial, the following is recommended:
a) Briefly centrifuge the tube
b) Open the cap and add 40 µl of TE buffer/ddH2O, tap gently
c) Do a second round of centrifugation to bring down contents

Q7: How do you typically transform plasmids?

A: Our plasmid transformation protocol is as follows:
a) Fetch one tube (size of 1.5 ml) containing 50 µl competent cells (stored at -80°C) and thaw it on ice for 20-30 min.
b) In the meanwhile dissolve the lyophilized 4 µg plasmid with sterilized water to make 100 ng/µl .
c) Mix 2~3 µl of plasmid with the competent cells by tapping the tube.
d) Place the tube on ice for 12-15 min.
e) Heat shock the tube by placing two-thirds of it into a 42°C water bath for 120 seconds.
f) Return the tube back on ice for 2 min.
g) Add 800 µl LB and shake in a 37°C incubator (200 rpm) for 40 min.
h) Plate some or all of the transformation mixture onto a LB agar plate containing antibiotic.
i) Incubate plates at 37°C overnight.

Q8: Do you have any suggestion on enzyme digestion upon receiving the plasmid?

A: The restriction enzymes we used in the COA file may not be present at the two ends of your insert. Basically, we use one or two unique and/or common restriction enzymes to identify and confirm that the insert has been successfully cloned in the vector of choice. Customer can specify the restriction enzymes to be used during QC while placing the order.

Q9: Sometimes the enzymes I choose do not digest the plasmid. Why?

A: If the recognition site is preceded by GA and followed by TC, or constitutes a CCWGG sequence (methylation site), you may not be able to digest the plasmid. Hence, it is recommended to avoid, if possible,enzymes such as: XbaI (TCTAGA), ClaI (ATCGAT), ApaI (GGGCCC), AvaII (GGWCC), SfiI (GGCCNNNNNGGCC), StuI (AGGCCT).

However, if you are unable to avoid using these enzymes on the methylated cloning sites in your plasmid, you may use a dam deficient strain such as GM2163 (E4105S), ER2925 or JM110 to amplify the DNA before digestion. You are requested to inform us of this while placing the order.

Q10: Why are some extra bases are added at cloning sites? Do these bases interfere with any downstream applications?

A: Protective bases are added outside the insert to facilitate cloning and normally should not interfere with your subcloning projects. For any special cloning manipulations,customers are requested to inform us, so we shall be able to work together to determine an ideal strategy.

Q11: Which strain is preferable for gRNA plasmid transformation?

A: Stbl3 strain is recommended.

Q12: How is the endotoxin (LAL) test performed?

A: The endotoxin test is a gel clot assay used to determine whether test samples contain specific amounts of endotoxin. The sample and the limulus amebocyte lysate (LAL) reagent solution are mixed in an endotoxin-free tube and incubated at 37°C for 60 min (see image A). After the incubation, the tube is tilted to see whether the mixture clots. If the sample contains less endotoxin than the standard, the solution remains liquid; whilst if the endotoxin amount is higher than the standard, the mixture will clot at the bottom of the tube (see image B).

A. Incubation at 37°C for 60 min
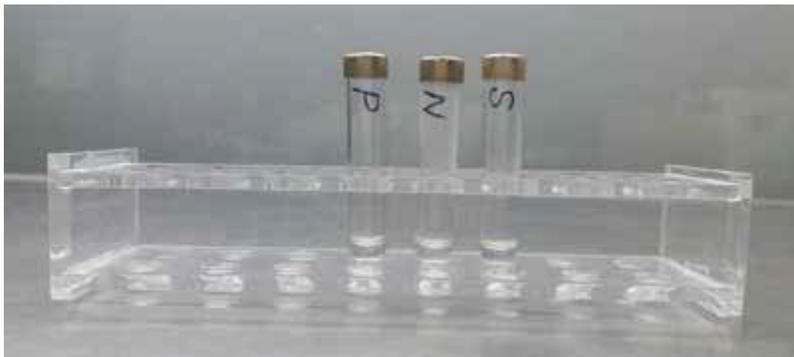
B. Endotoxin Test
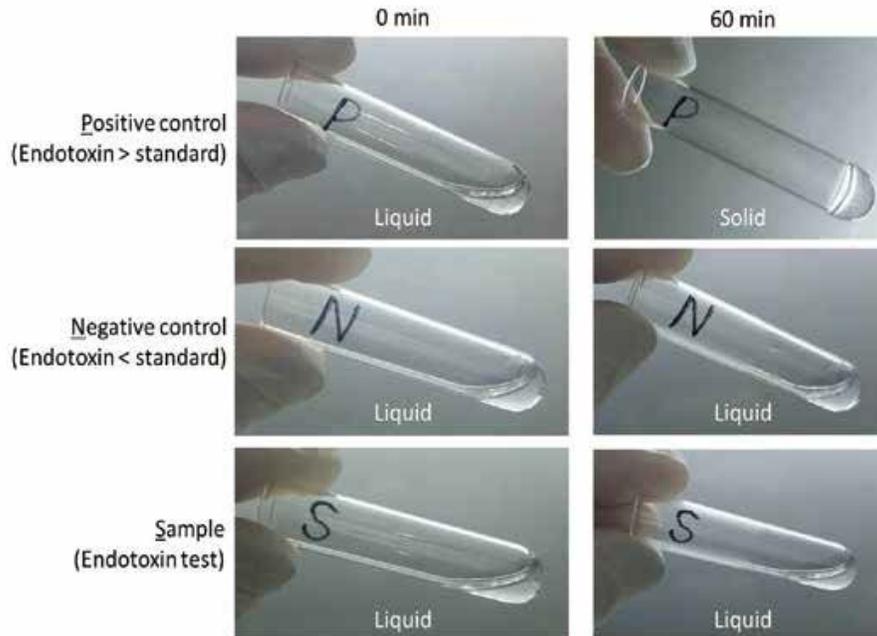
Q13: What materials/equipment are required to make a Luria broth (LB) Agar Plates

- 37 g pre-mixed powder consisting of:
    - 5 g peptone
    - 10.0 g peptone from casein
    - 10.0 g sodium chloride
    - 12.0 g agar-agar
- 1 L Sterile H_2O
- Sterile plates of your desired size - we usually use 60 mm x 15 mm plates which can hold 5-10 mL of agar and on which you can individually distinguish a maximum of ~100 colonies
- Autoclavable flasks
- Sterile pipettes
- Ice bucket to hold antibiotic
- Antibiotic at 1000 x concentration disolved in the appropriate liquid solvent. See table to the right for appropriate antibiotic concentrations.

Q14: What is the standard Antibiotic Concentrations?

Antibiotic	Recommended Stock Concentration	Recommended Working Concentration
Ampicillin	100 mg/mL	100 µg/mL
Bleocin	5 mg/mL	5 µg/mL

Antibiotic	Recommended Stock Concentration	Recommended Working Concentration
Carbenicillin*	100 mg/mL	100 µg/mL
Chloramphenicol	25 mg/mL (dissolve in EtOH)	25 µg/mL
Coumermycin	25 mg/mL (dissolve in DMSO)	25 µg/mL
Gentamycin	10 mg/mL	10 µg/mL
Kanamycin	50 mg/mL	50 µg/mL
Spectinomycin	50 mg/mL	50 µg/mL
Tetracycline	10 mg/mL	10 µg/mL

Note: Unless otherwise indicated, the antibiotic powder can be dissolved in dH_2O.

Note: *Carbenicillin can be used in place of ampicillin. Carbenicillin is more stable, so it is potentially more effective at selecting only bacteria containing the plasmids of interest (for example, fewer satellite colonies will grow). However, it is also more expensive.*

Q15: If I have any questions about your service (product/data), how can I contact you?

A: You may always contact: support@novoprolabs.com
Or the scientific consultant you were contacting with the orders/inquiries.

Supporting References (from Addgene)

Plasmid Elements

Plasmids used by scientists today come in many sizes and vary broadly in their functionality. In their simplest form, plasmids require a bacterial origin of replication (*ori*), an antibiotic-resistance gene, and at least one unique restriction enzyme recognition site. These elements allow for the propagation of the plasmid within

bacteria, while allowing for selection against any bacteria not carrying the plasmid. Additionally, the restriction enzyme site(s) allow for the cloning of a fragment of DNA to be studied into the plasmid.

Below are some common plasmid elements:

Plasmid Element	Description
Origin of Replication (*ori*)	DNA sequence which directs initiation of plasmid replication (by bacteria) by recruiting DNA replication machinery. The *ori* is critical for the ability of the plasmid to be copied (amplified) by bacteria, which is an important characteristic of why plasmids are convenient and easy to use.
Antibiotic Resistance Gene	Allows for selection of plasmid-containing bacteria by providing a survival advantage to the bacterial host. Each bacterium can contain multiple copies of an individual plasmid, and ideally would replicate these plasmids upon cell division in addition to their own genomic DNA. Because of this additional replication burden, the rate of bacterial cell division is reduced (i.e., it takes more time to copy this extra DNA). Because of this reduced fitness, bacteria without plasmids can replicate faster and out-populate bacteria with plasmids, thus selecting against the propagation of these plasmids through cell division. To ensure the retention of plasmid DNA in bacterial populations, an antibiotic resistance gene (i.e., a gene whose product confers resistance to ampicillin) is included in the plasmid. These bacteria are then grown in the presence of ampicillin. Under these conditions, there is a selective pressure to retain the plasmid DNA, despite the added replication burden, as bacteria without the plasmid DNA would not survive antibiotic treatment. It is important to distinguish that the antibiotic resistance gene is under the control of a bacterial promoter, and is thus expressed in the bacteria by bacterial transcriptional machinery.
Multiple Cloning Site (MCS)	Short segment of DNA which contains several restriction enzyme sites, enabling easy insertion of DNA by restriction enzymes digestion and ligation. In expression plasmids, the MCS is often located downstream from a promoter, such that when a gene is inserted within the MCS, its expression will be driven by the promoter. As a general rule, the restriction sites in the MCS are unique and not located elsewhere in the plasmid backbone, which is why they can be used for cloning by

Plasmid Element	Description
	restriction enzyme digestion. For more information about restriction enzymes check out NEB's website .
Insert	The insert is the gene, promoter, or other DNA fragment cloned into the MCS. The insert is typically the genetic element one wishes to study using a particular plasmid.
Promoter Region	Drives transcription of the insert. The promoter is designed to recruit transcriptional machinery from a particular organism or group of organisms. Meaning, if a plasmid in intended for use in human cells, the promoter will be a human or mammalian promoter sequence. The promoter can also direct cell-specific expression, which can be achieved by a tissue-specific promoter (e.g., a liver-specific promoter). The strength of the promoter is also important for controlling the level of insert expression (i.e., a strong promoter directs high expression, whereas weaker promoters can direct low/endogenous expression levels). For more information about promoters, both bacterial and eukaryotic, as well as common promoters used in research check out our promoters reference page.
Selectable Marker	The selectable marker is used to select for cells that have successfully taken up the plasmid for the purpose of expressing the insert. This is different than selecting for bacterial cells that have taken up the plasmid for the purpose of replication. The selectable marker enables selection of a population of cells that have taken up the plasmid and that can be used to study the insert. The selectable marker is typically in the form of another antibiotic resistance gene (this time, under the control of a non-bacterial promoter) or a fluorescent protein (that can be used to select or sort the cells by visualization or FACS).
Primer Binding Site	A short single-stranded DNA sequence used as an initiation point for PCR amplification or DNA sequencing of the plasmid. Primers can be utilized to verify the sequence of the insert or other regions of the plasmid. For commonly used primers check out Addgene's sequencing primer list.

Working with Plasmids

Plasmids have become an essential tool in molecular biology for a variety of reasons, including that they are:

- **Easy to work with** - Plasmids are a convenient size (generally 1,000-20,000 basepairs) for physical isolation (purification) and manipulation. With current cloning technology, it is easy to create and modify plasmids containing the genetic element that you are interested in.

- **Self-replicating** - Once you have constructed a plasmid, you can easily make an endless number of copies of the plasmid using bacteria, which can uptake plasmids and amplify them during cell division. Because bacteria are easy to grow in a lab, divide relatively quickly, and exhibit exponential growth rates, plasmids can be replicated easily and efficiently in a laboratory setting.

- **Stable** - Plasmids are stable long-term either as purified DNA or within bacterial cells that have been preserved as glycerol stocks.

- **Functional in many species and can be useful for a diverse set of applications** - Plasmids can drive gene expression in a wide variety of organisms, including plants, worms, mice, and even cultured human cells. Although plasmids were originally used to understand protein coding gene function, they are now used for a variety of studies used to investigate promoters, small RNAs, or other genetic elements.

Types of Plasmids

Plasmids are versitile and can be used in many different ways by scientists. The combination of elements often determines the type of plasmid and dictates how it might be used in the lab. Below are some common plasmid types:

- **Cloning Plasmids** - Used to facilitate the cloning of DNA fragments. Cloning vectors tend to be very simple, often containing only a bacterial resistance gene, origin of replication, and MCS. They are small and optimized to help in the initial cloning of a DNA fragment. Commonly used cloning vectors include Gateway entry vectors and TOPO cloning vectors. *If you are looking for an empty plasmid backbone for your experiment, see* Addgene's empty backbone page *for more information.*

- **Expression Plasmids** - Used for gene expression (for the purposes of gene study). Expression vectors must contain a promoter sequence, a transcription terminator sequence, and the inserted gene. The promoter region is required for

the generation of RNA from the insert DNA via transcription. The terminator sequence on the newly synthesized RNA signals for the transcription process to stop. An expression vector can also include an enhancer sequence which increases the amount of protein or RNA produced. Expression vectors can drive expression in various cell types (mammalian, yeast, bacterial, etc.), depending largely on which promoter is used to initiate transcription.

- **Gene Knock-down Plasmids** - Used for reducing the expression of an endogenous gene. This is frequently accomplished through expression of an shRNA targeting the mRNA of the gene of interest. These plasmids have promoters that can drive expression of short RNAs.

- **Genome Engineering Plasmids** - Used to target and edit genomes. Genome editing is most commonly accomplished using CRISPR technology. CRISPR is composed of a DNA endonuclease and guide RNAs that target specific locations in the genome. For more information on CRISPR check out Addgene's CRISPR guide.

- **Reporter Plasmids** - Used for studying the function of genetic elements. These plasmids contain a reporter gene (for example, luciferase or GFP) that offers a read-out of the activity of the genetic element. For instance, a promoter of interest could be inserted upstream of the luciferase gene to determine the level of transcription driven by that promoter.

- **Viral Plasmids** - These plasmids are modified viral genomes that are used to efficiently deliver genetic material into target cells. You can use these plasmids to create viral particles, such as lentiviral, retroviral, AAV, or adenoviral particles, that can infect your target cells at a high efficiency. *Addgene's expanding viral service offers select ready-made AAV and lentiviral particles. Visit our viral service page to learn more.*

Regardless of type, plasmids are generally propagated, selected for, and the integrity verified prior to use in an experiment.

E. coli strains for propagating plasmids

E. coli are gram-negative, rod shaped bacteria naturally found in the intestinal tract of animals. There are many different naturally occurring strains of *E. coli*, some of which are deadly to humans. The majority of all common, commercial lab strains of *E. coli* used today are descended from two individual isolates, the K-12 strain and the B strain. K-12 has led to the common lab strains MG1655 and its derivatives DH5alpha and DH10b (also known as TOP10) among others, while the B strain gave rise to BL21 and its derivatives.

We've included a small number of *E. coli* strains below and recommend checking out these two Addgene blog posts relating to common *E. coli* lab strains and *E. coli* strains

specialized for protein expression for additional strain-related information and a more extensive strain list.

Strain	Vendor(s)	Genotype
BL21	Invitrogen; New England BioLabs	*E. coli* B F dcm ompT hsdS(rB mB) gal
ccdB Survival	Invitrogen	F- mcrA Delta(mrr-hsdRMS-mcrBC) Phi80lacZDeltaM15 Delta-lacX74 recA1 araDelta139 D(ara-leu)7697 galU galK rpsL (StrR) endA1 nupG tonA::Ptrc ccdA
DB3.1	Invitrogen	F- gyrA462 endA Delta(sr1-recA) mcrB mrr hsdS20 (rB-mB-) supE44 ara14 galK2 lacY1 proA2 rpsL20(StrR) xyl5 lambda- leu mtl1
DH5alpha	Invitrogen	F- Phi80lacZDeltaM15 Delta(lacZYA-argF) U169 recA1 endA1 hsdR17(rk-, mk+) phoA supE44 thi-1 gyrA96 relA1 tonA
JM109	Addgene; Promega	e14-(McrA-) recA1 endA1 gyrA96 thi-1 hsdR17(rK-mK+) supE44 relA1 Delta(lac- proAB) [F traDelta36 proAB lacIqZDeltaM15]
NEB Stable	New England Biolabs	F' proA+B+ lacIq Δ(lacZ)M15 zzf::Tn10 (TetR) Δ(ara-leu) 7697 araD139 fhuA ΔlacX74 galK16 galE15 e14-Φ80dlacZΔM15 recA1 relA1 endA1 nupG rpsL (StrR) rph spoT1 Δ(mrr-hsdRMS-mcrBC)
Stbl3	Invitrogen	F– mcrB mrr hsdS20 (rB–, mB–) recA13 supE44 ara-14 galK2 lacY1 proA2 rpsL20 (StrR) xyl-5 λ– leu mtl-1
Top10	Invitrogen	F- mcrA Delta(mrr-hsdRMS-mcrBC) Phi80lacZM15 Delta-lacX74 recA1 araD139 Delta(ara-leu)7697 galU galK rpsL (StrR) endA1 nupG

Antibiotics commonly used for plasmid selection

Many plasmids are designed to include an antibiotic resistance gene, which when expressed, allows only plasmid-containing bacteria to grow in or on media containing that antibiotic. These antibiotic resistance genes not only give the scientist with an easy way to detect plasmid-containing bacteria, but also provide those bacteria with a pressure to maintain and replicate your plasmid over multiple generations. More information relating to antibiotic resistance genes as well as additional antibiotics not listed in the table below can be found in this blog post.

Below you will find a few antibiotics commonly used in the lab and their recommended concentrations. We suggest checking your plasmid's datasheet or the plasmid map to confirm which antibiotic(s) to add to your LB media or LB agar plates.

Antibiotic	Recommended Stock Concentration	Recommended Working Concentration
Ampicillin	100 mg/mL	100 μg/mL
Carbenicillin*	100 mg/mL	100 μg/mL
Chloramphenicol	25 mg/mL (dissolve in EtOH)	25 μg/mL
Hygromycin B	200 mg/mL	200 μg/mL
Kanamycin	50 mg/mL	50 μg/mL
Spectinomycin	50 mg/mL	50 μg/mL
Tetracycline	10 mg/mL	10 μg/mL

Note: Carbenicillin can be used in place of ampicillin.

Preparing Antibiotics

1. Create a stock solution of your antibiotic. Unless otherwise indicated, the antibiotic powder can be dissolved in dH_2O. Addgene recommends making 1000X stock solutions and storing aliquots at -20°C.

2. To use, dilute your antibiotic into your LB medium at 1:1,000. For example, to make 100 mL of LB/ampicillin growth media, add 100 µL of a 100 mg/mL ampicillin stock (1000X stock) to 100 mL of LB.

DNA sequencing for plasmid verification

DNA is made up of 4 bases, adenine, thymine , cytosine, and guanine. The order of these bases makes up the genetic code and provides all the information needed for cells to make proteins and other molecules essential for life.

Scientists often "sequence DNA" to identify the order of these four nucleotide bases in a particular DNA strand. Sequencing DNA and understanding the genetic code allows scientists to study gene function as well as identify changes or mutations that may cause certain diseases. **Sequencing DNA is extremely important when verifying plasmids to ensure each plasmid contains the essential elements to function and the correct gene of interest.** So how do scientists sequence DNA?

Sanger Sequencing

In 1975, Frederick Sanger developed the process termed Sanger sequencing, sometimes referred to as chain-termination sequencing or dideoxy sequencing.

To understand Sanger sequencing, we first need to understand DNA replication. DNA is a double helix, where a base on one strand pairs with a particular base on the other, complementary, strand. Specifically, A pairs with T and C pairs with G. During replication, DNA unwinds and the DNA polymerase enzyme binds to and migrates down the single stranded DNA adding nucleotides according to the sequence of the complementary strand.

The replication process can also be done in a test tube to copy DNA regions of interest. In vitro DNA replication requires the 4 nucleotides, a DNA polymerase enzyme, the template DNA to be copied, and a primer. A primer is a small piece of DNA, approximately 18-22 nucleotides, that binds to complementary DNA and acts as a starting point for the DNA polymerase. Thus to replicate a piece of DNA *in vitro* one has to know some of its sequence to design a effective primer.

Sanger sequencing is modeled after in vitro DNA replication but relies on the random incorporation of modified, fluorescently tagged bases onto the growing DNA strand in addition to the normal A, T, C, or G nucleotide. The 4 standard bases are tagged with a different fluorophore so they can be distinguished from one another. Similar to DNA replication, the Sanger sequencing reaction begins when a primer binds to its complementary DNA and the DNA polymerase adding nucleotides. The major difference

in this process occurs when the polymerase incorporates a fluorescently tagged nucleotide. Because these special bases do not have a binding site for adding the next nucleotide, the reaction is halted once the fluorescently tagged base is incorporated.

Sanger sequencing requires a lot of DNA because the ultimate goal is to have a fluorescently tagged nucleotide at each position in the DNA sequence. Thus, the final result is a group of newly synthesized DNA strands of varying lengths whose last nucleotide is labeled. Once all the newly synthesized DNA is made, the DNA molecules are then separated by size from shortest to longest and "read" using a sequencing machine that recognizes the different fluorescent labels. The machine detects which fluorescently labeled nucleotide is present at the end of each fragment and assembles that information into the DNA sequence. Sanger sequencing results are presented as a sequencing chromatogram which provides the color and intensity of each fluorescent signal. Sanger can sequence approximately 500-1000 bases downstream of the known primer region with very few errors making it an efficient and reliable sequencing method.

Next Generation Sequencing

Although Sanger sequencing is quick and efficient, it is low throughput and can only sequence short pieces of DNA. This is not extremely useful when trying to sequence an entire plasmid or an organism's genome. One Sanger sequencing reaction would give you only 20 pieces of a 2,000 piece puzzle. A scientist would need to run a ton of Sanger sequencing reactions on different pieces of DNA to be able to assemble the whole puzzle. That's where Next Generation Sequencing (NGS) comes in. NGS is a high-throughput, multi parallel sequencing platform that can generate sequencing data for up to 600 billion bases in one reaction. In other words NGS can give you most of the puzzle pieces in only a few reactions.There are multiple approaches to acquire NGS but one of the most commonly used is the Illumina NGS platform. This is the platform used by Addgene's sequencing partner, seqWell.

The actual process of Illumina NGS is not that different from Sanger sequencing. This process, like Sanger, is based on DNA replication and utilizes modified fluorescently tagged nucleotides. During illumina NGS, a long piece of DNA is first fragmented into small pieces, labeled with a short DNA barcode, and amplified. These DNA fragments are attached to a glass slide so that different fragments of DNA, or templates, are spatially separated from each other. These attached DNA templates are then amplified again producing ~1,000 copies of each template. Each template is then replicated using the modified bases and a microscope captures the fluorescent color that is emitted each time a base is added. Again, each base (A,C,T, or G) is labelled with a different color making it easy to identify the order of the DNA strand. Unlike Sanger however, these modified bases can be converted back to a regular base and thus do not halt the reaction. Illumina NGS, therefore does not require any "normal" bases in the reaction. All the sequenced templates are then aligned to each other to assemble the entire sequence or puzzle. It is important to note that NGS platforms in general do not require

a specific primer for your DNA of interest thus a completely unknown piece of DNA can be sequenced.

At Addgene all incoming plasmids are sequenced with NGS during our quality control process. NGS allows us to sequence entire plasmids providing scientists with even more information to aid in the reproducibility of scientific research.

Resources

Genetic Code

The genetic code can be defined as a set of rules for translating the information encoded by DNA and RNA into proteins. DNA is comprised of 4 nucleotides Adenine (A), Thymine (T), Cytosine (C) and Guanine (G). In the double helix A always pairs with T and C always pairs with G. RNA on the other hand consists of Adenine, Cytosine, Guanine and Uracil (U). Uracil replaces thymine in RNA molecules. Every 3 nucleotides (codons) in a DNA sequence encodes for an amino acid. The genetic code is degenerate thus multiple codons code for each amino acid. There are 20 amino acids plus a start and stop codon.

Below you will find helpful resource tables about the genetic code. This table includes the nucleotide and amino acid code in addition to ambiguous bases and common epitope tags. Ambiguous bases are included in a DNA sequence when sequencing is not 100% efficient and the machine cannot distinguish between the 4 labelled nucleotides. Epitope tags on the other hand are commonly used in molecular cloning to tag a gene within a plasmid.

DNA and RNA

Single Letter Code: Primary bases	Nucleobase
A	Adenine
C	Cytosine
G	Guanine
T	Thymine

Single Letter Code: Primary bases	Nucleobase
U	Uracil
B	C, G, or T
D	A, G, or T
H	A, C, or T
K	G or T
M	A or C
N	A, T, C, or G
R	A or G
S	C or G
V	A, C, or G
W	A or T
Y	C or T

Amino Acids

Name	Three Letter Code	Single Letter Code	Codons (RNA)
Alanine	Ala	A	GCU, GCC, GCA, GCG
Arginine	Arg	R	CGU, CGC, CGA, CGG, AGA, AGG
Asparagine	Asn	N	AAU, AAC
Aspartic acid	Asp	D	GAU, GAC
Cysteine	Cys	C	UGU, UGC
Glutamine	Gln	Q	CAA, CAG
Glutamic Acid	Glu	E	GAA, GAG
Glycine	Gly	G	GGU, GGC, GGA, GGG
Histidine	His	H	CAU, CAC
Isoleucine	Ile	I	AUU, AUC, AUA
Leucine	Leu	L	UUA, UUG, CUU, CUC, CUA, CUG
Lysine	Lys	K	AAA, AAG
Methionine	Met	M	AUG

Name	Three Letter Code	Single Letter Code	Codons (RNA)
Phenylalanine	Phe	F	UUU, UUC
Proline	Pro	P	CCU, CCC, CCA, CCG
Serine	Ser	S	UCU, UCC, UCA, UCG, AGU,AGC
Threonine	Thr	T	ACU, ACC, ACA, ACG
Tryptophan	Trp	W	UGG
Tyrosine	Tyr	Y	UAU, UAC
Valine	Val	V	GUU, GUC, GUA, GUG
Start			AUG*
Stop			UAG (amber), UGA (opal), UAA (ochre)

*AUG is the most common start codon. Alternative start codons include CUG in eukaryotes and GUG in prokaryotes.

Common Epitope Tags

Tag	Amino Acid Sequence
FLAG	DYKDDDDK
HA	YPYDVPDYA

Tag	Amino Acid Sequence
His	HHHHHH
Myc	EQKLISEEDL
V5	GKPIPNPLLGLDST
Xpress	DLDDDDK or DLYDDDDK
Thrombin	LVPRGS
BAD (Biotin Acceptor Domain)	GLNDIFEAQKIEWHE
Factor Xa	IEGR or IDGR
VSVG	YTDIEMNRLGK
SV40 NLS	PKKKRKV or PKKKRKVG
Protein C	EDQVDPRLIDGK
S Tag	KETAAAKFERQHMDS
SB1	PRPSNKRLQQ

Origins of Molecular Genetics

The concept of genes as carriers of phenotypic information was introduced in the early 19th century by Gregor Mendel, who later demonstrated the properties of genetic inheritance in peas. Over the next 100 years, many significant discoveries lead to the conclusions that genes encode proteins and reside on chromosomes, which are

composed of DNA. These findings culminated in the central dogma of molecular biology, that proteins are translated from RNA, which is transcribed from DNA.

DNA is comprised of 4 nucleotides or bases, adenine, thymine, cytosine, and guanine (abbreviated to A, T, C, and G respectively) that are organized into a double stranded helix. The order of these 4 nucleotides makes up the genetic code and provides the instructions to make every protein within an organism. Proteins are made up of amino acids. Each amino acid is encoded for by 3 nucleotides termed a codon. As there are only 20 natural amino acids and 64 codon combinations each amino acid is encoded for by multiple codons.

Plasmids and Recombinant DNA Technology

Techniques in chemistry enable isolation and purification of cellular components, such as DNA, but practically this isolation is only feasible for relatively short DNA molecules. In order to isolate a particular gene from human chromosomal DNA, it would be necessary to isolate a sequence of a few hundred or few thousand basepairs from the entire human genome. Digesting the human genome with restriction enzymes would yield about two million DNA fragments, which is far too many to separate from each other for the purposes of isolating one specific DNA sequence. This obstacle has been overcome by the field of recombinant DNA technology, which enables the preparation of more managable (i.e., smaller) DNA fragments.

In 1952, Joshua Lederberg coined the term plasmid, in reference to any extrachromosomal heritable determinant. Plasmids are fragments of double-stranded DNA that typically carry genes and can replicate independently from chromosomal DNA. Although they can be found in archaea and eukaryotes, they play the most significant biological role in bacteria where they can be passed from one bacterium to another by a type of horizontal gene transfer (conjugation), usually providing a benefit to the host, such as antibiotic resistance. This benefit can be context-dependent, and thus the plasmid exists in a symbiotic relationship with the host cell. Like the bacterial chromosomal DNA, plasmid DNA is replicated upon cell division, and each daughter cell receives at least one copy of the plasmid.

By the 1970s the combined discoveries of restriction enzymes, DNA ligase, and gel electrophoresis allowed for the ability to move specific fragments of DNA from one context to another, such as from a chromosome to a plasmid. These tools are essential to the field of recombinant DNA, in which many identical DNA fragments can be generated. The combination of a DNA *fragment* with a plasmid or vector DNA backbone generates a *recombinant* DNA molecule, which can be used to study DNA fragments of interest, such as genes.

Creation of recombinant DNA

Molecular Cloning

Plasmids that are used most commonly in the field of recombinant DNA technology have been optimized for their use of studying and manipulating genes. For instance, most plasmids are replicated in *E. coli* and are relatively small (~3000 - 6000 basepairs) to enable easy manipulation. Typically plasmids contain the minimum essential DNA sequences for this purpose, which includes a DNA replication origin, an antibiotic-resistance gene, and a region in which exogenous DNA fragments can be inserted. When a plasmid exists extrachromosomally in *E. coli*, it is replicated independently and segregated to the resulting daughter cells. These daughter cells contain the same genetic information as the parental cell, and are thus termed *clones* of the original cell. The plasmid DNA is similarly referred to as *cloned* DNA, and this process of generating multiple identical copies of a recombinant DNA molecule is known as *DNA* or *molecular cloning*. The process of molecular cloning enabled scientists to break chromosomes down to study their genes, marking the birth of molecular genetics.

Today, scientists can easily study and manipulate genes and other genetic elements using specifically engineered plasmids, commonly referred to as vectors, which have become possibly the most ubiquitous tools in the molecular biologist's toolbox. To learn more about different types of cloning methods check out our guide on molecular cloning techniques.