UNIT IX

PART A

Site-directed Mutagenesis (SDM)

GEB, Sem VI, B.Sc (H) Biochemistry
Dr. Jayita Thakur

Michael Smith, a British-born Canadian biochemist won the **1993 Nobel Prize** (with Kary B. Mullis) in **Chemistry** for the development of a technique called -

oligonucleotide-based site-directed mutagenesis, which enabled researchers to introduce specific mutations into genes and, thus, to the proteins that they encode.

Site-directed mutagenesis (SDM) is a method to create specific, targeted changes in double stranded plasmid DNA.

There are many reasons to make specific DNA alterations (insertions, deletions and substitutions), including:

- To study changes in protein activity that occur as a result of the DNA manipulation.

- To select or screen for mutations (at the DNA, RNA or protein level) that have a desired property.

- To introduce or remove restriction endonuclease sites or tags.

Using site-directed mutagenesis,

- scientists have been able to dissect the structure and function relationships involved in protein plaque formation in the pathophysiology of Alzheimer disease;
- study the feasibility of gene therapy approaches for cystic fibrosis, sickle-cell disease, and hemophilia;
- determine the characteristics of protein receptors at neurotransmitter binding sites and design analogs with novel pharmaceutical properties;
- examine the viral proteins involved in immunodeficiency disease;
- and improve the properties of industrial enzymes used in food science and technology.

STRATEGIES OF DIRECTED MUTAGENESIS

A. Oligonucleotide-Directed Mutagenesis with M13 DNA
B. Oligonucleotide-Directed Mutagenesis with Plasmid DNA
C. PCR based mutagenesis
    i. PCR-Amplified Oligonucleotide-Directed Mutagenesis
    ii. Error-Prone PCR
    iii. Random Mutagenesis with Degenerate Oligonucleotide Primers
D. Random Insertion/Deletion Mutagenesis
E. DNA Shuffling
F. Mutant Proteins with Unusual Amino Acids

A. Oligonucleotide-directed mutagenesis with M13 DNA

Single-stranded bacteriophage M13 (M13 + strand), carrying a cloned gene is taken.

↓

It is annealed with a complementary synthetic oligonucleotide containing one mismatched base.

↓

With the oligonucleotide as the primer, DNA synthesis is catalyzed by the Klenow fragment of *E. coli* DNA polymerase I.

↓

Synthesis continues until the entire strand is copied. The newly synthesized DNA strand is circularized by T4 DNA ligase.

↓

The ligation reaction mixture is used to transform *E. coli*.

↓

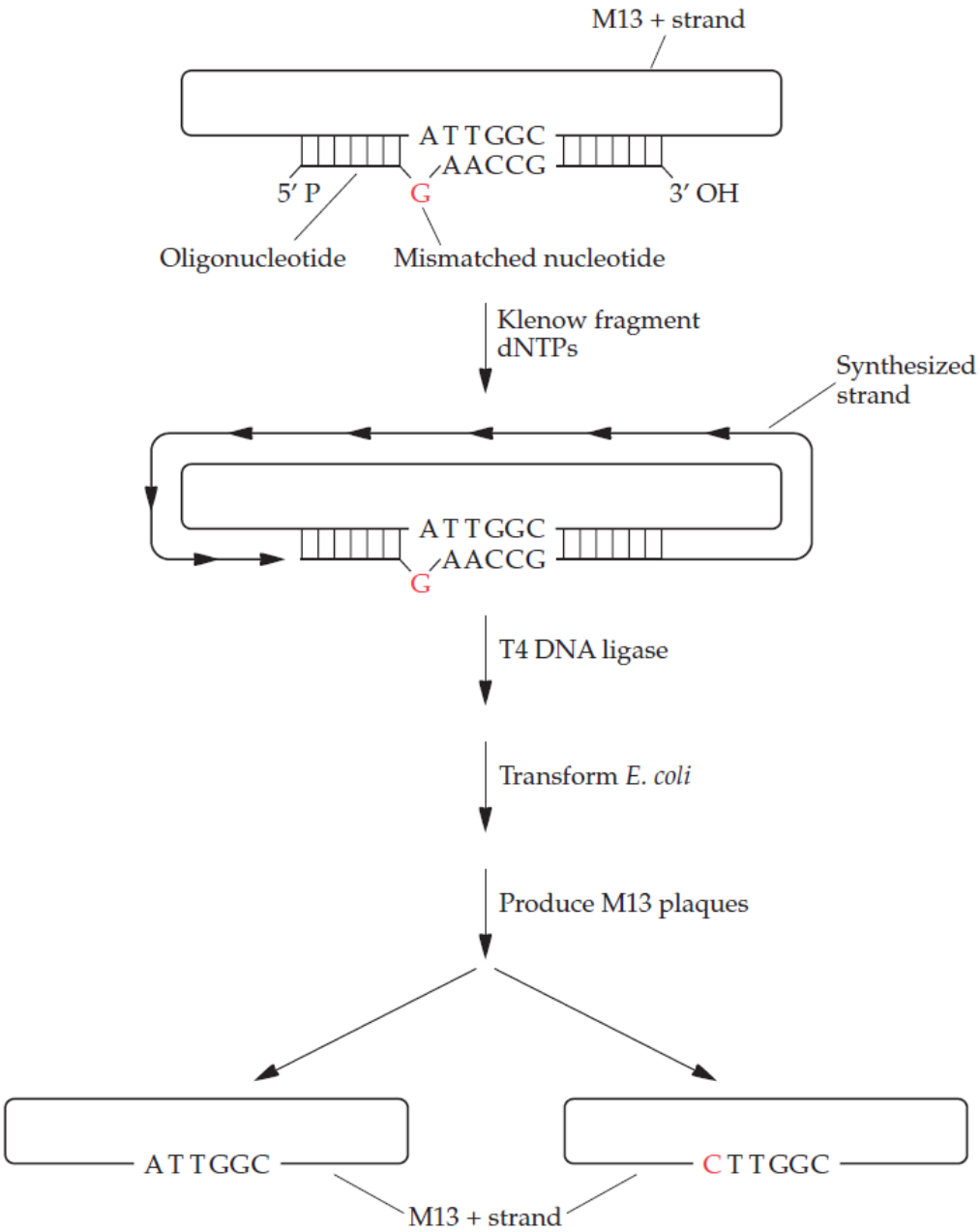
Both the target DNA with its original sequence and the mutated sequence are present in the progeny M13 phage.

↓

After the double-stranded form of M13 is isolated, the mutated gene is excised by digestion with restriction enzymes and then spliced onto an *E. coli* plasmid expression vector.

↓

For further study, the altered protein is expressed in and purified from the *E. coli* cells.


M13 + strand
ATTGGC
AACCG
5′ P
G
3′ OH
Oligonucleotide
Mismatched nucleotide
Klenow fragment
dNTPs
Synthesized
strand
ATTGGC
AACCG
G
T4 DNA ligase
Transform E. coli
Produce M13 plaques
ATTGGC
CTTGGC
M13 + strand

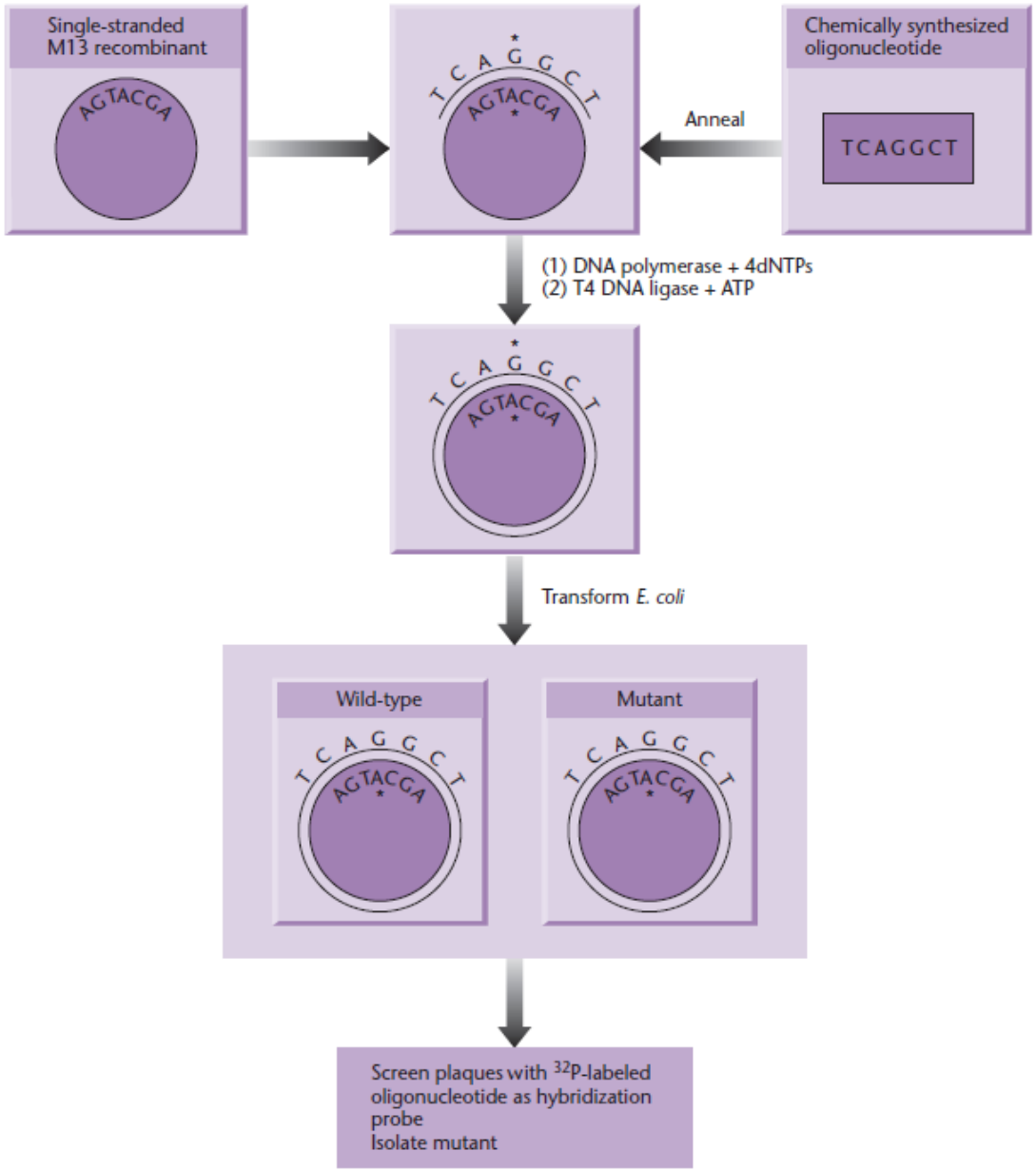


Fig. 8.1
Oligonucleotide-directed mutagenesis. Asterisks indicate mismatched bases. Originally the Klenow fragment of DNA polymerase was used, but now this has been largely replaced with T7 polymerase.

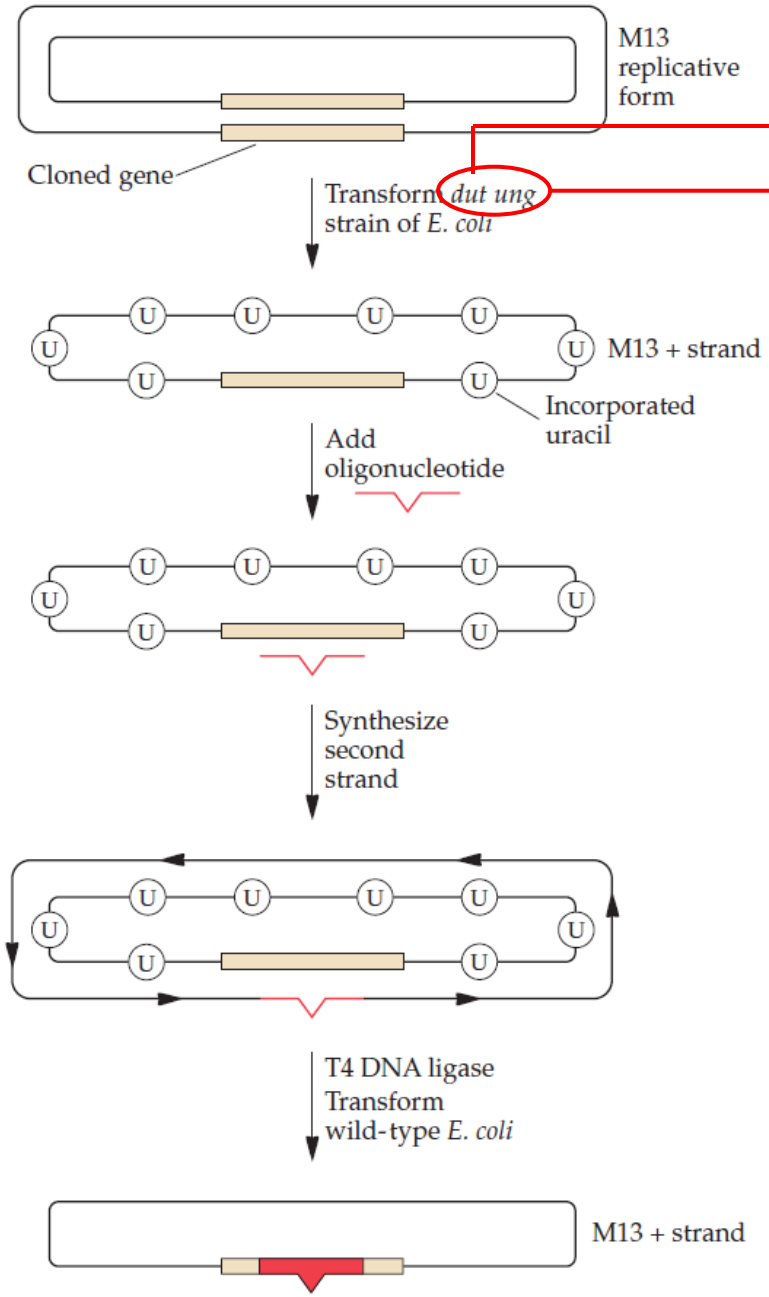
As per the diagram we see only 50% of the M13 viruses are expected to carrying the mutated form of the target.

In practice only around 1% of the plaques actually contain phage carrying the mutated gene.

Consequently, the oligonucleotide-directed mutagenesis method has been modified in several ways to enrich for the number of mutant phage plaques that can be obtained.

Modification

Introduce the M13 viral vector carrying the gene that is to be mutagenized into an *E. coli* strain that has two defective enzymes of DNA metabolism.



dUTPase *(dut)*

Cells without a functional dUTPase have an elevated intracellular level of dUTP, which in turn causes a few dUTP residues to be incorporated into DNA during replication instead of dTTP.

uracil *N*-glycosylase *(ung)*

In the absence of functional uracil *N*-glycosylase, the dUTP residues that were spuriously incorporated into DNA cannot be removed.

The target DNA is cloned into the double-stranded replicative form of bacteriophage M13, which is then used to transform a *dut ung* strain of *E. coli*.

↓

The *dut* mutation causes the intracellular level of dUTP to be elevated; the high level of nucleotide leads to the incorporation of a few dUTP residues (U). The *ung* mutation prevents the removal of any incorporated uracil residues.

↓

Following in vitro oligonucleotide-directed mutagenesis, the double-stranded M13 vector with the mutated DNA is introduced into wild-type *E. coli*.

↓

The wild-type *ung* gene product removes any uracil residues from the parental strand, so a significant portion of the parental strand is degraded.

↓

The mutated strand remains intact because it does not contain uracil. It serves as a template for DNA replication, thereby enriching the yield of M13 bacteriophage carrying the mutated gene.

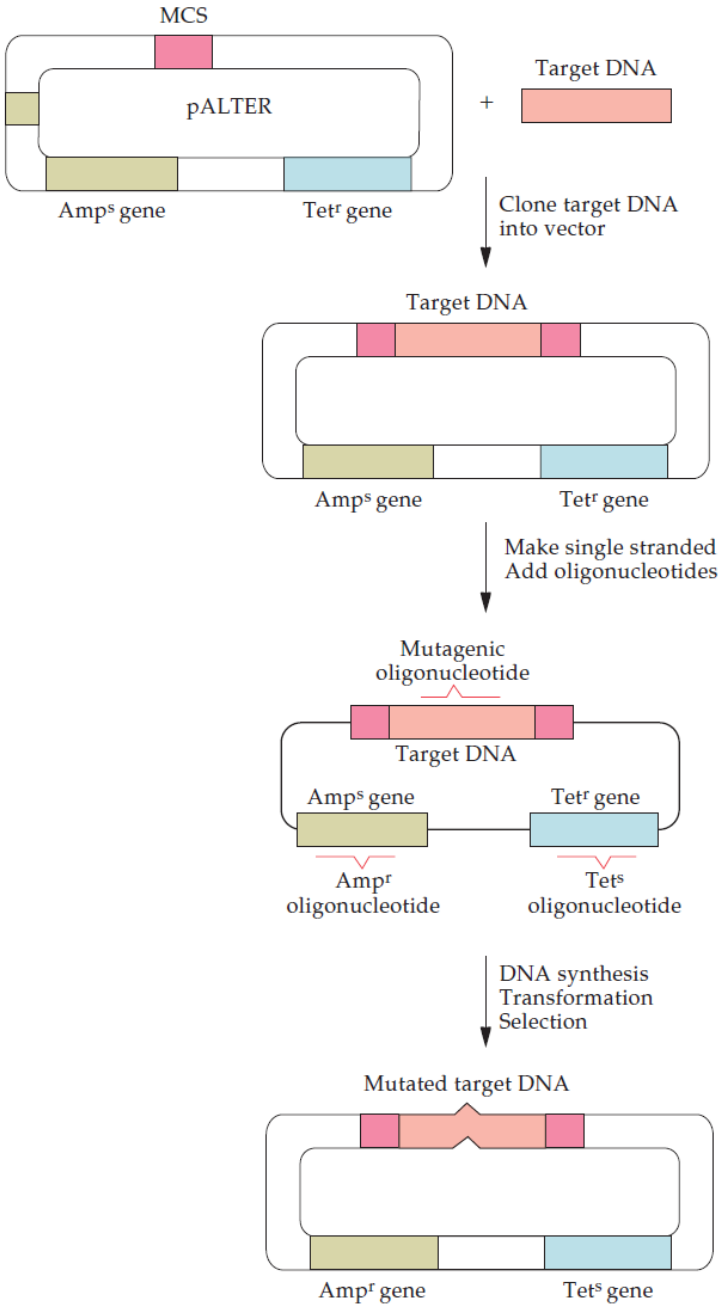
Drawbacks of using the M13 DNA for carrying out SDM

- There is a need to subclone a target gene from a plasmid into M13 and then, after mutagenesis, clone it back into a plasmid.

- Additional step of transforming enzyme defective *E. coli* necessary to enrich the yield.

- Lengthy process involving multiple steps

B. Oligonucleotide-Directed Mutagenesis with Plasmid DNA



The target DNA is inserted into the multiple cloning site (MCS) on the vector pALTER.

Plasmid DNA is isolated from *E. coli* cells and alkaline denatured.

Mutagenic oligonucleotide, the ampicillin resistance (Ampr) oligonucleotide, and the tetracycline sensitivity (Tets) oligonucleotide are annealed.

The oligonucleotides act as primers for DNA synthesis by T4 DNA polymerase with the original strand as the template.

The gaps between the synthesized pieces of DNA are sealed by T4 DNA ligase.
The reaction mixture is used to transform *E. coli* host cells, and cells that are Ampr and Tets are selected.

- With this procedure, about 90% of the selected transformants have the specified mutation in the target gene.

- In the remaining transformants, the target gene is unchanged because the oligonucleotide did not anneal to the target gene or it was bypassed during DNA synthesis.

- The cells with the specified mutation in the target gene are identified by DNA hybridization.

- All of the plasmids, host bacterial strains, enzymes, oligonucleotides (other than the one needed to alter the target gene), and buffers for this method are sold as a kit, facilitating its widespread use.

C. PCR based mutagenesis

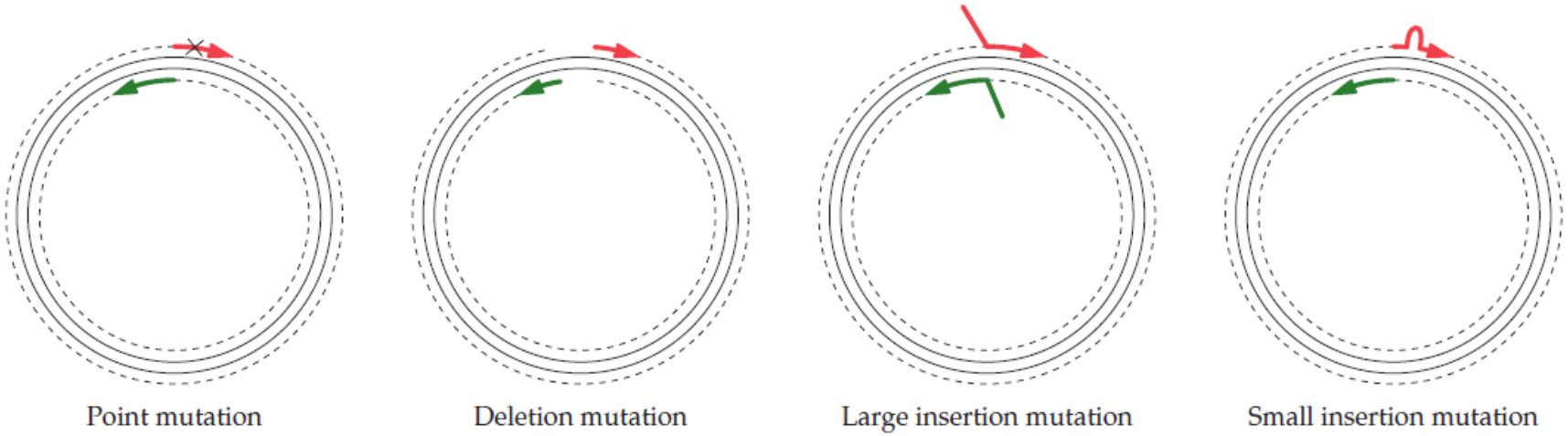
C. (i)PCR-Amplified Oligonucleotide-Directed Mutagenesis

- For this method of mutagenesis, no special plasmid vectors are required; any plasmid up to approximately 10 kb in length is acceptable.

- For PCR-based mutagenesis point mutations, nucleotide changes are introduced in the middle of the primer sequence.

- To create deletion mutations, primers must border the region of target DNA to be deleted on both sides and be perfectly matched to their annealing (or template) sequences.

- To create mutations with long insertions, a stretch of mismatched nucleotides is added to the 5' end of one or both primers, while for mutations with short insertions, a stretch of nucleotides is designed in the middle of one of the primers.



Overview of the basic methodology to introduce point mutations, insertions, or deletions into DNA cloned into a plasmid.

In all of these procedures, the only absolute requirements are that:
(1) the nucleotide sequence of the target DNA must be known.
(2) the 5' ends of the primers must be phosphorylated.

Following PCR amplification, the linear DNA is circularized by ligation with T4 DNA ligase.

The circularized plasmid DNA is then used to transform *E. coli* by any standard procedure.

This protocol yields a very high frequency of plasmids with the desired mutation. The screening three or four clones by sequencing the target DNA should be sufficient to find the desired

In summary, this procedure introduces a specified mutation (point, insertion, or deletion) into a cloned gene without the need to insert the cloned gene into bacteriophage M13; to use enrichment procedures, such as the *dut ung* system; or to subclone the mutated gene from M13 onto an expression plasmid vector.

Given its simplicity and effectiveness, this procedure has come to be widely used.

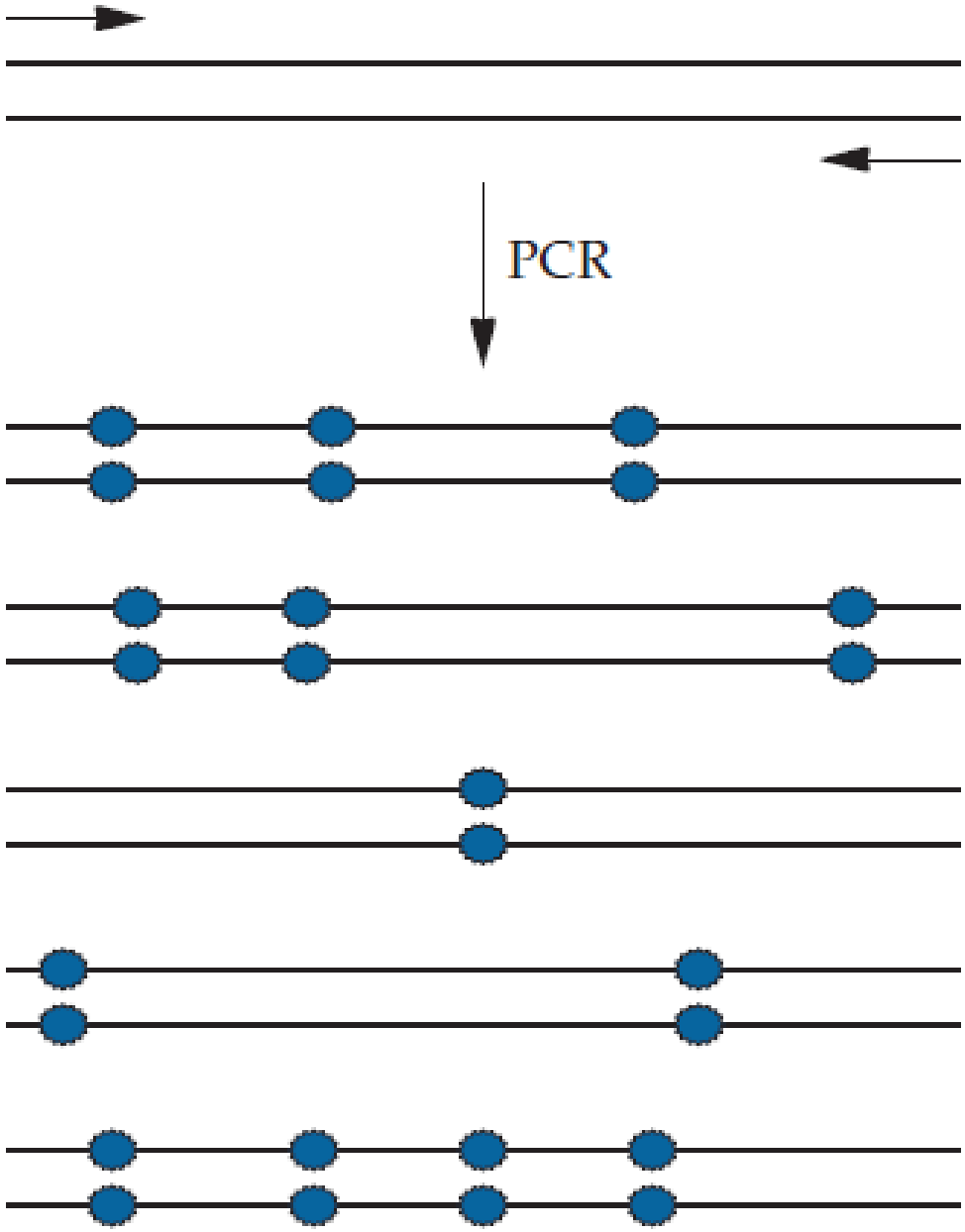
C. (ii) Error-Prone PCR

This is a very powerful method for **random** mutagenesis useful for the **construction of a library of mutants**.

With DNA up to 10 kb in size, it is possible to vary the number of alterations per gene from about 1 to about 20 by modifying the DNA template concentration.

When error-prone PCR is performed using *Taq* DNA polymerase, which lacks proofreading activity, the error rate may be increased-

A. by adding Mn2+, by increasing the concentration of Mg2+
B. by adding unequal amounts of the four deoxynucleoside triphosphates to the reaction buffer.
C. by using other temperature-stable DNA polymerases in the absence of Mn2+ and with balanced amounts of the four deoxynucleoside triphosphates.


PCR

- Following error-prone PCR, the randomly mutagenized DNA is cloned into expression vectors and screened for altered or improved protein activity.

- The DNA from those clones that encode the desired activity is isolated and sequenced so that the relevant changes to the target DNA may be elaborated.

- Error-prone PCR has been used to create enzymes with improved solvent and temperature stability and with enhanced specific activity.

Drawback

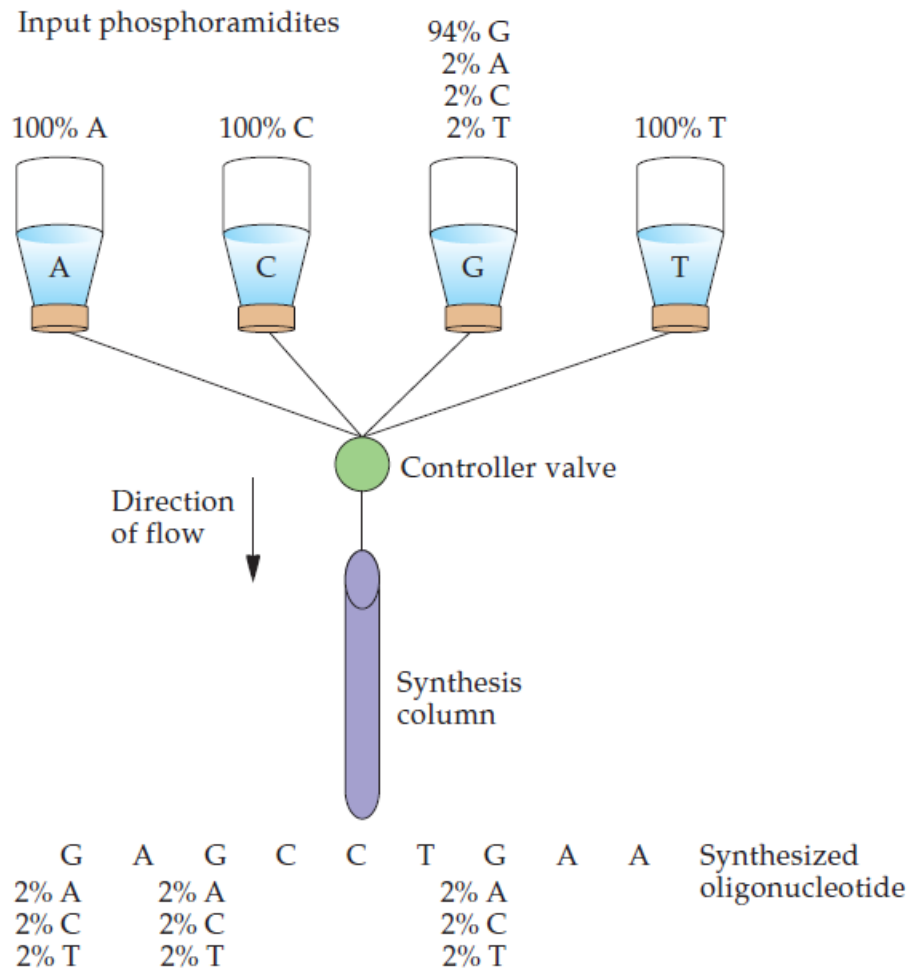
Since errors are typically introduced into DNA at no more than one or two per 1,000 nucleotides, only single nucleotides are replaced within a triplet codon, yielding only a limited number of amino acid changes.

C. (iii) Random Mutagenesis with Degenerate Oligonucleotide Primers

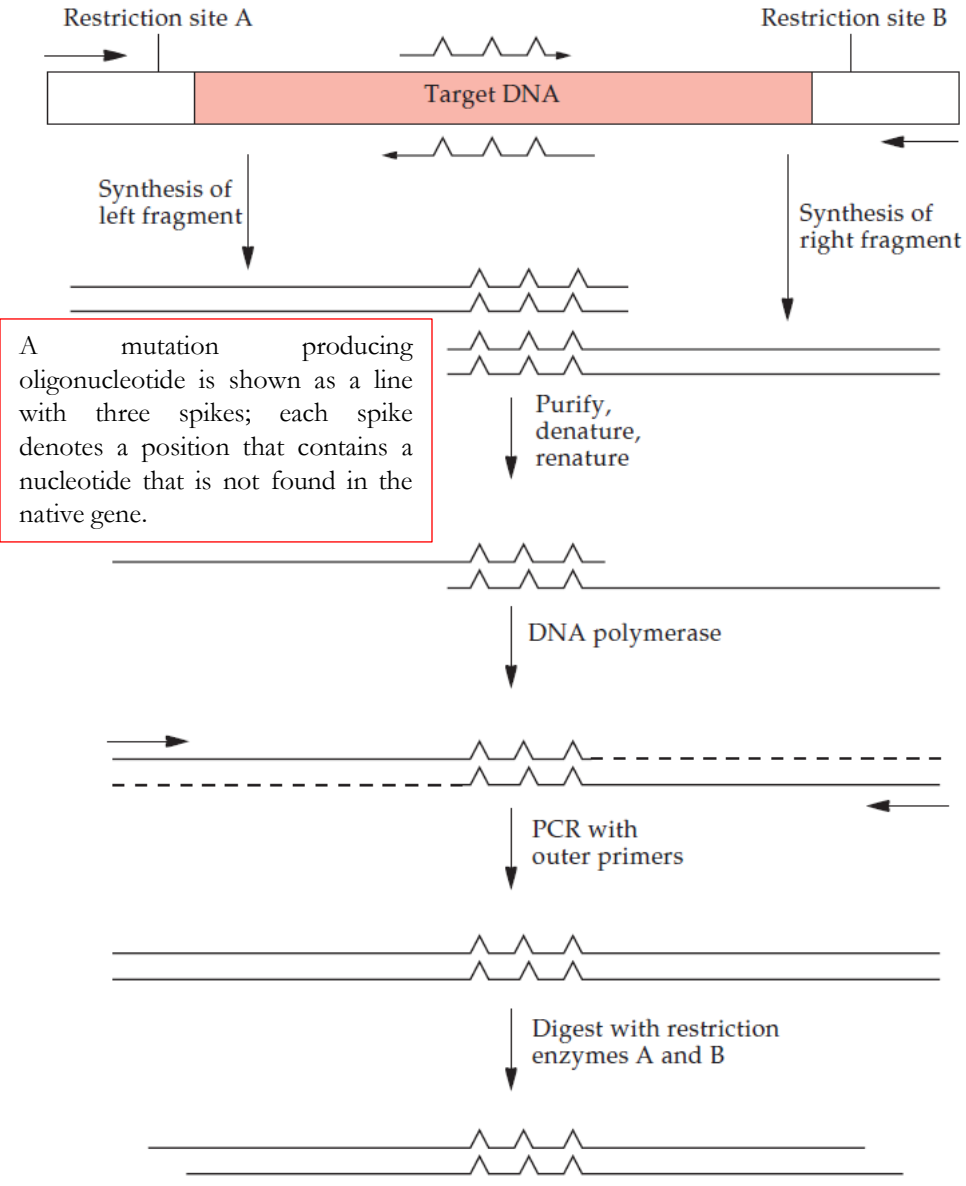
Investigators seldom know which specific nucleotide changes need to be introduced into a cloned gene to modify the properties of the target protein.

Consequently, they must use methods that generate all the possible amino acid changes at one particular site.

For example, oligonucleotide primers can be synthesized with any of the four nucleotides at defined positions. This pattern of sequence degeneracy is generally achieved by programming an automated DNA synthesis reaction to add a low level (usually a few percent) of each of the three alternative nucleotides each time a particular nucleotide is added to the chain.



In this way, the oligonucleotide primer preparation contains a heterogeneous set of DNA sequences that will generate a series of mutations that are clustered in a defined portion of the target gene.



A target gene is inserted into a plasmid between two unique restriction endonuclease sites.

The left and right portions of the target DNA are amplified separately by PCR.
The amplified fragments are purified, denatured to make them single stranded, and then reannealed.

Complementary regions of overlap are formed between complementary mutation producing oligonucleotides. The single-stranded regions are made double stranded with DNA polymerase, and then the entire fragment is amplified by PCR.

The resultant product is digested with restriction endonucleases A and B and then cloned into a vector that has been digested with the same enzymes.

This approach has two advantages.

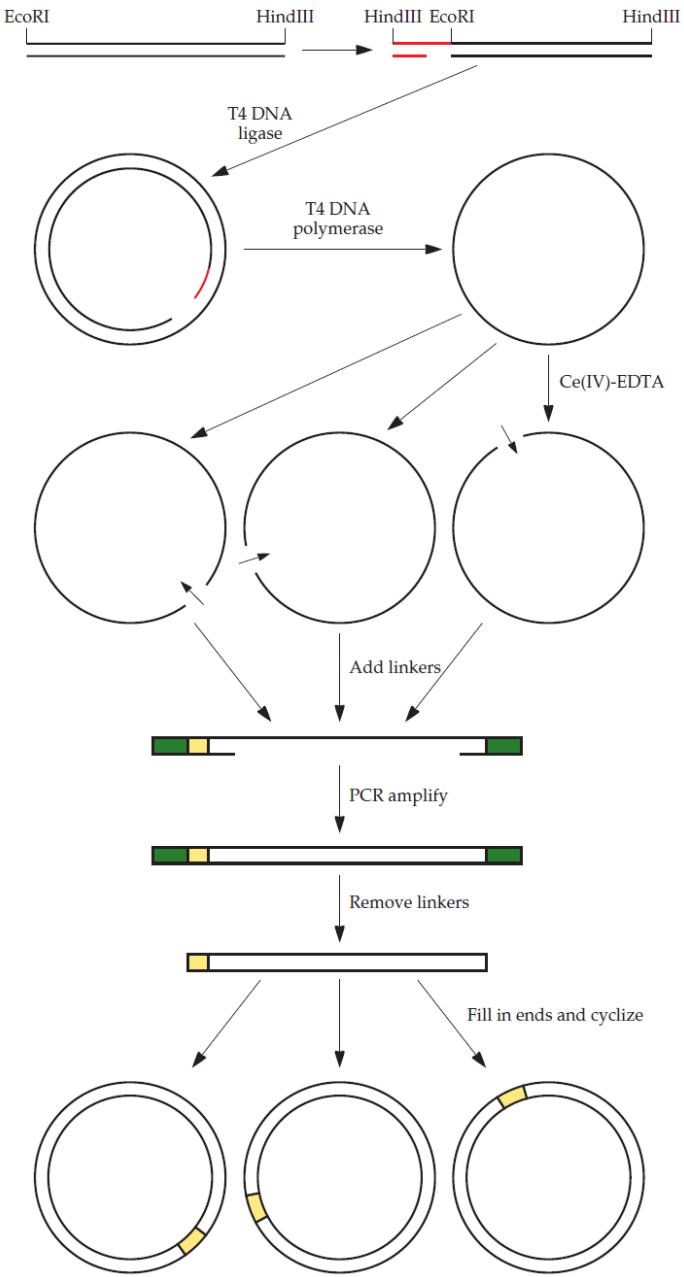
(1) Detailed information regarding the roles of particular amino acid residues in the functioning of the protein is not required.

(2) Unexpected mutants encoding proteins with a range of interesting and useful properties may be generated because the introduced changes are not limited to one amino acid.

D. Random Insertion/Deletion Mutagenesis

As an alternative to error-prone PCR, researchers have developed the technique of random insertion/deletion mutagenesis.

With this approach, it is possible to delete a small number of nucleotides at random positions along the gene and, at the same time, insert either specific or random sequences into that position.


EcoRI
HindIII
HindIII EcoRI
HindIII
T4 DNA ligase
T4 DNA polymerase
Ce(IV)-EDTA
Add linkers
PCR amplify
Remove linkers
Fill in ends and cyclize

1. An isolated gene fragment with different restriction endonuclease sites at each end is ligated at one end to a short nonphosphorylated linker that leaves a small gap in the DNA. The gap is a consequence of the fact that the 5′ nucleotide from the linker is not phosphorylated and therefore cannot be ligated to an adjacent 3′-OH group.

2. After restriction enzyme digestion that creates compatible sticky ends, the gene fragment is cyclized with T4 DNA ligase to create a circular double-stranded gene fragment with a nick in the antisense strand.

3. The nicked strand is degraded by digestion with the enzyme T4 DNA polymerase (which has exonuclease activity).

4. The single-stranded DNA is randomly cleaved at single positions by treating it with a cerium(IV)–ethylenediaminetetraacetic acid (EDTA) complex.

5. The linear single-stranded DNAs are ligated to a linker (containing several additional nucleotides selected for insertion at one end), and the entire mutagenesis library is PCR amplified.

6. The linkers are removed by restriction enzyme digestion.

7. The constructs are made blunt ended by filling in the single stranded overhangs using the Klenow fragment of *E. coli* DNA polymerase I and then cyclized again by T4 DNA ligase.

8. The amplified products are digested with appropriate restriction enzymes, cloned into a plasmid vector, and then tested for activity.

With this approach, it is possible to insert any small DNA fragment (carried on a linker) into the randomly cleaved single-stranded DNA, with the result that a much greater number of modified genes may be generated than by error-prone PCR.

The mutations that are developed by this procedure may be used to select protein variants with a wide range of activities.

E. DNA Shuffling

Wild type

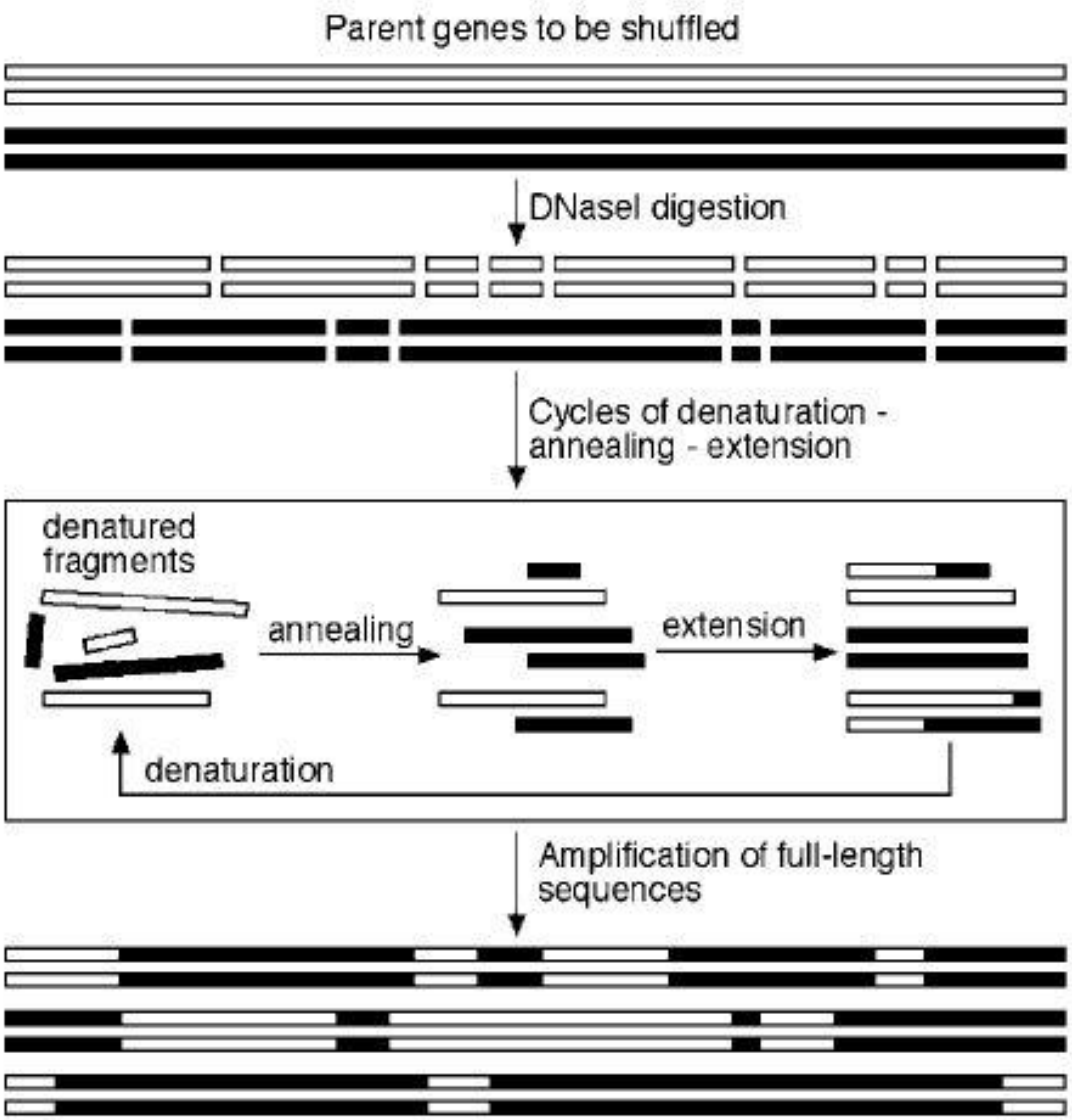
Random
mutagenesis
or error-prone PCR

DNA shuffling

DNA shuffling is a method for in vitro recombination of homologous genes invented by W.P.C Stemmer.

It is done in the hope that some of the hybrid proteins will have unique properties or activities that were not encoded in any of the original sequences.

Also, some of the hybrid proteins may combine important attributes of two or more of the original proteins, e.g., high activity and thermostability.



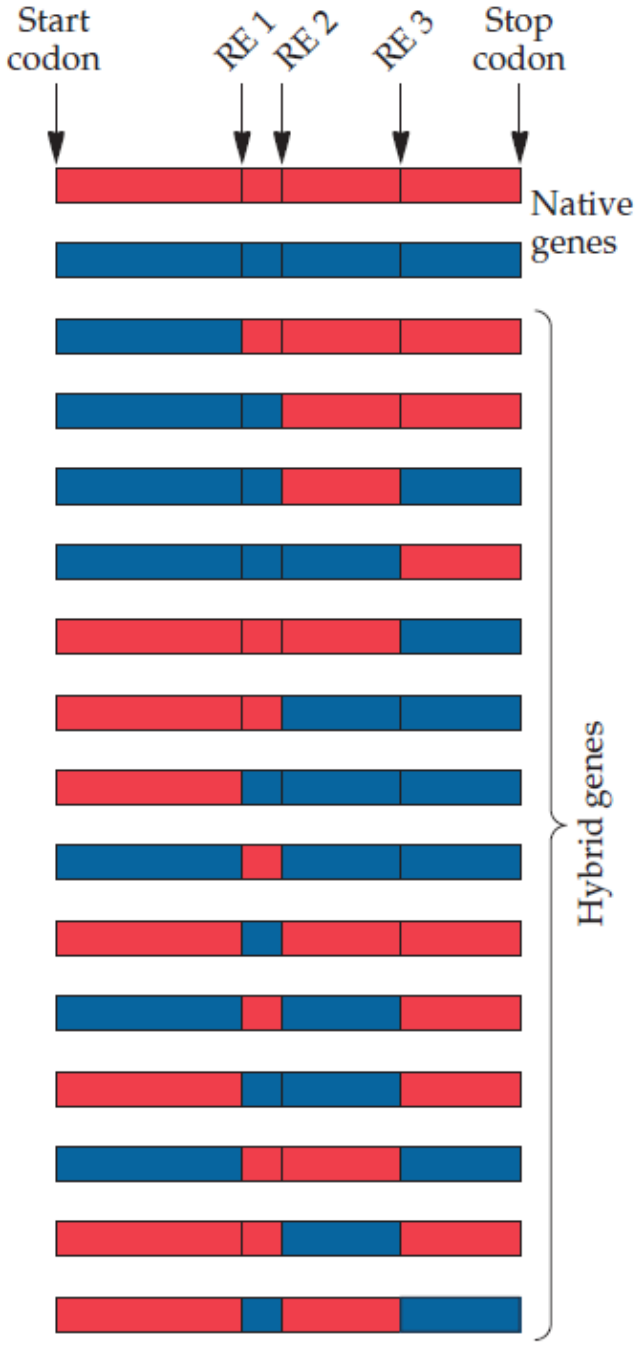
The genes to be recombined are randomly fragmented by DNaseI, and fragments of the desired size are purified from an agarose gel.

These fragments are then reassembled using cycles of denaturation, annealing, and extension by a polymerase.

Recombination occurs when fragments from different parents anneal at a region of high sequence identity.

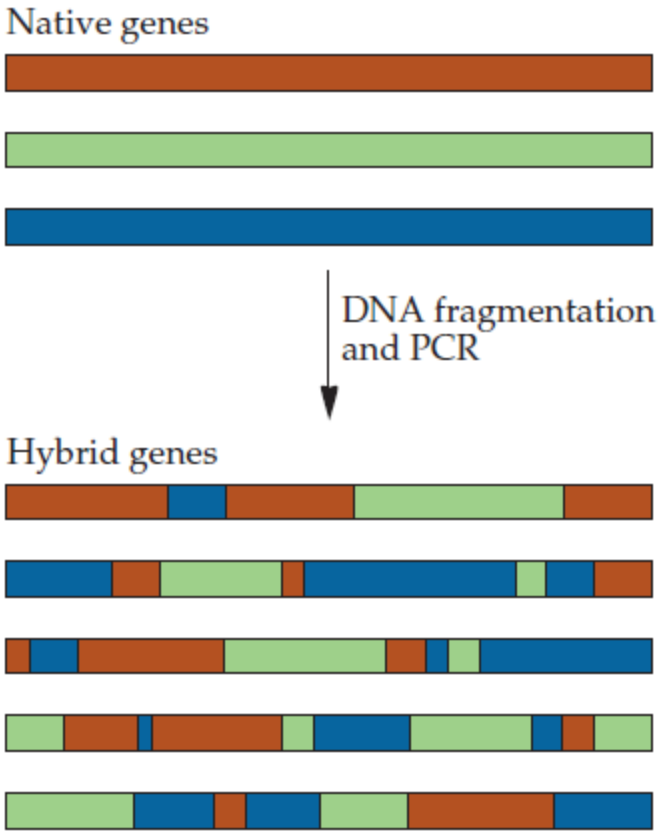
Following this reassembly reaction, PCR amplification with primers is used to generate full-length chimeras suitable for cloning into an expression vector.

https://link.springer.com/protocol/10.1385/1-59259-395-X:85



The simplest way to shuffle portions of similar genes is through the use of common restriction enzyme sites.

Digestion of two or more of the DNAs that encode the native forms of similar proteins with one or more restriction enzymes that cut the DNAs in the same place, followed by ligation of the mixture of DNA fragments, can potentially generate a large number of hybrids.



Another way to shuffle DNA involves

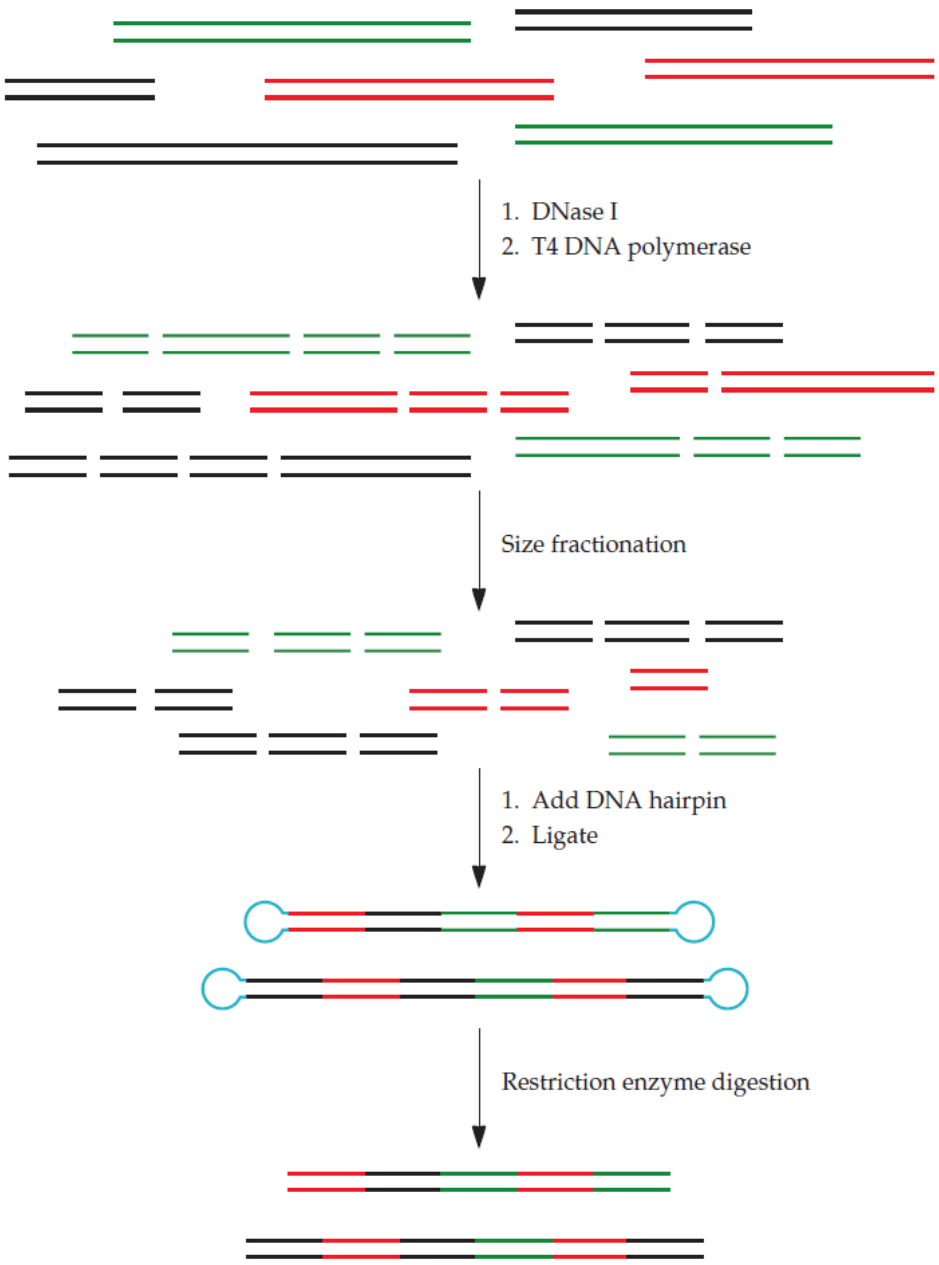
- Combining several members of a gene family, fragmenting the mixed DNA with deoxyribonuclease I (DNase I), selecting smaller DNA fragments
- Followed by PCR amplifying these fragments.
- During PCR, gene fragments from different members of a gene family cross-prime each other after DNA fragments bind to one another in regions of high homology/complementarity.

- The final full-length products are obtained by including "terminal primers" in the PCR. After 20 to 30 PCR cycles, a panel of hybrid (full-length) DNAs will be established.
- The hybrid DNAs are then used to create a library that can be screened for the desired activity.

Although DNA shuffling works well with gene families—it is sometimes called molecular breeding—or with genes from different families that nevertheless have a high degree of homology, the technique is not especially useful when proteins have little or no homology.

Thus, the DNAs must be very similar to one another or the PCR will not proceed.

To remedy this situation and combine the genes of dissimilar proteins, several variations of the DNA-shuffling protocol have been developed.



Different DNAs (shown in different colors) are mixed together, partially digested with DNase I, blunt ended by digestion with T4 DNA polymerase.

They are then size fractionated.

The fragments are ligated with synthetic hairpin DNAs to form extended hairpins.

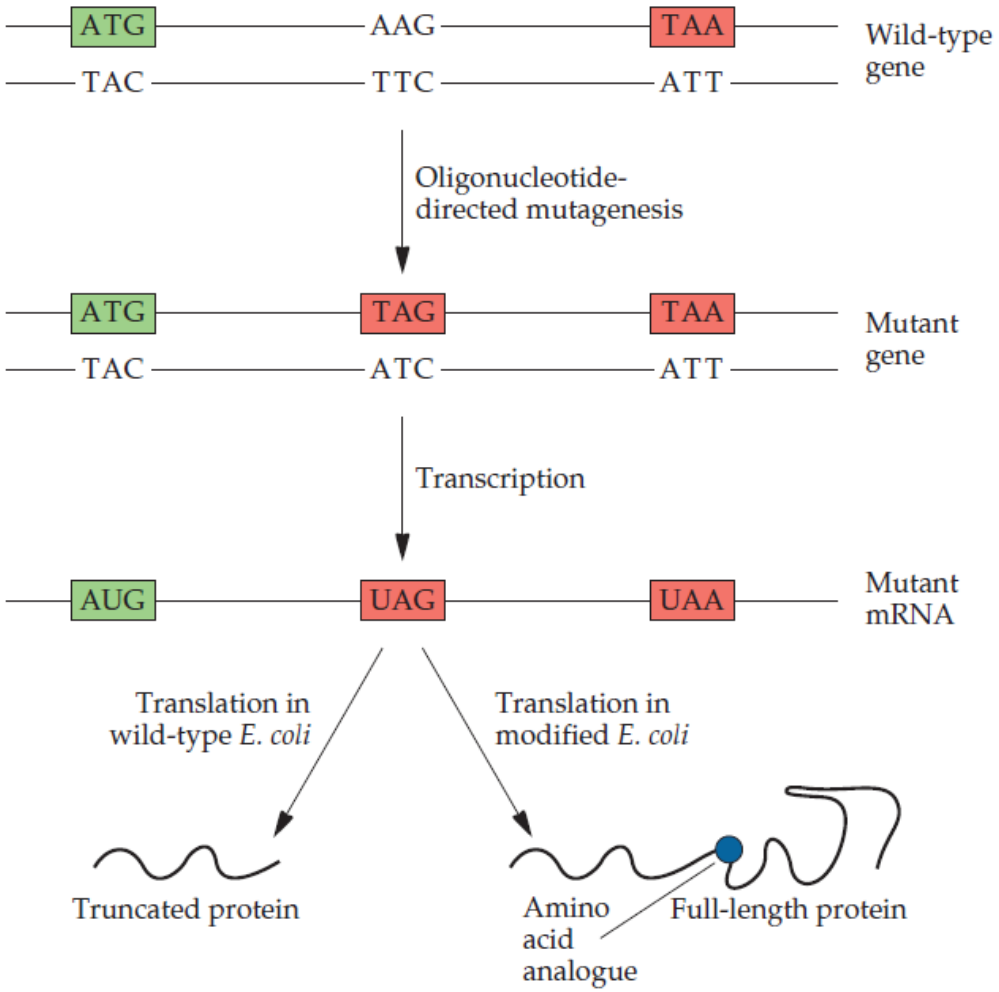
Restriction enzymes digest and remove the hairpin ends and generate sticky ends, and then ligated into plasmid vectors.

F. Mutant Proteins with Unusual Amino Acids

Any protein can be altered by substituting one amino acid for another using directed mutagenesis.

There are only 20 amino acids that are normally used in protein synthesis, hence this becomes a limitation in generating mutants

One way to increase the diversity of the proteins formed after mutagenesis is to introduce synthetic amino acids with unique side chains at specific sites. To do this, *E. coli* was engineered to produce both a novel transfer RNA (tRNA) that is not recognized by any of the existing *E. coli* aminoacyl-tRNA synthetases but nevertheless functions in translation and a new aminoacyltRNA synthetase that aminoacylates only that novel tRNA.



Schematic representation of the production of a protein with a modified (nonstandard amino acid) side chain. The start codon is highlighted in green, and the stop codons are in red. The inserted amino acid analogue is shown in blue.