Department of BSBE
Indian Institute Of Technology Guwahati

Introduction to microbiology and study of microorganisms:

In-frame protein fusion and use of GFP for subcellular localization of protein/DNA

Dr. Sanjukta Patra

BT 207

Genetic Engineering

Jan-April 2023

Fusion protein are chimeric proteins.

It is a protein consisting of at least two domains that are encoded by separate genes.

They have been joined so that they are transcribed and translated as a single unit, producing a single polypeptide.

They have a linker in between.

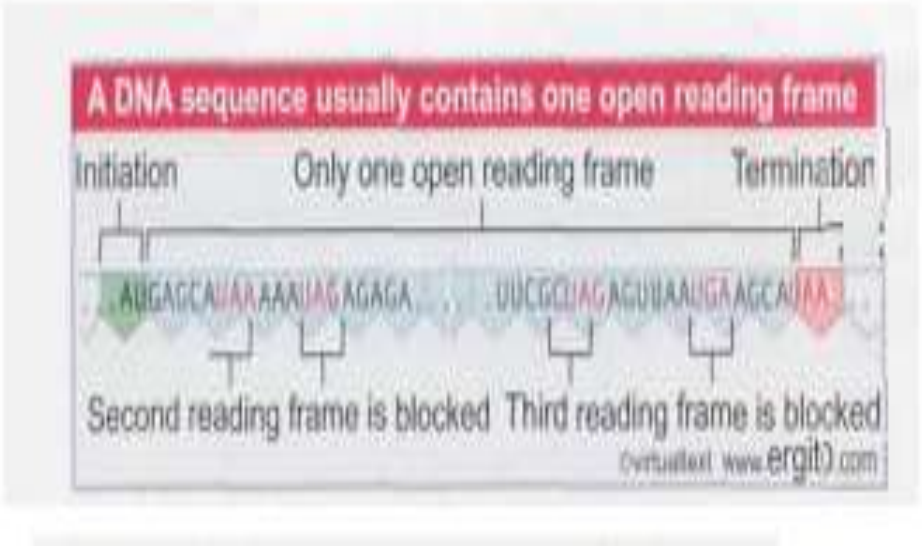
Fusion proteins can be created in vivo, for example, as the result of a chromosomal rearrangement.

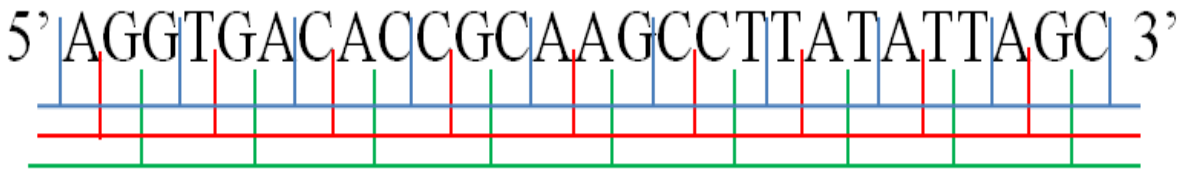
In-frame protein fusion and use of GFP for subcellular localization of protein/DNA

- **In frame** - fusion in frame" means the fusion of two proteins/peptides in the same orientation/direction,
- so that the cloned product will have an ORF containing both protein sequences.
- Otherwise there may be a frameshift, which may truncate further events of expression.
- This indicates that two proteins are expressed together

Every sequence has three possible reading frames

- A **reading frame** is a way of dividing the sequence of nucleotides in a nucleic acid (DNA or RNA) molecule into a set of consecutive, non-overlapping triplets.
- Where these triplets equate to amino acids or stop signals during translation.
- A single strand of a nucleic acid molecule has a phosphoryl end, called the 5′-end, and a hydroxyl or 3′-end. These define the 5′→3′ direction.
- There are three reading frames that can be read in this 5′→3′ direction, each beginning from a different nucleotide in a triplet.
- Usually only one reading frame is translated



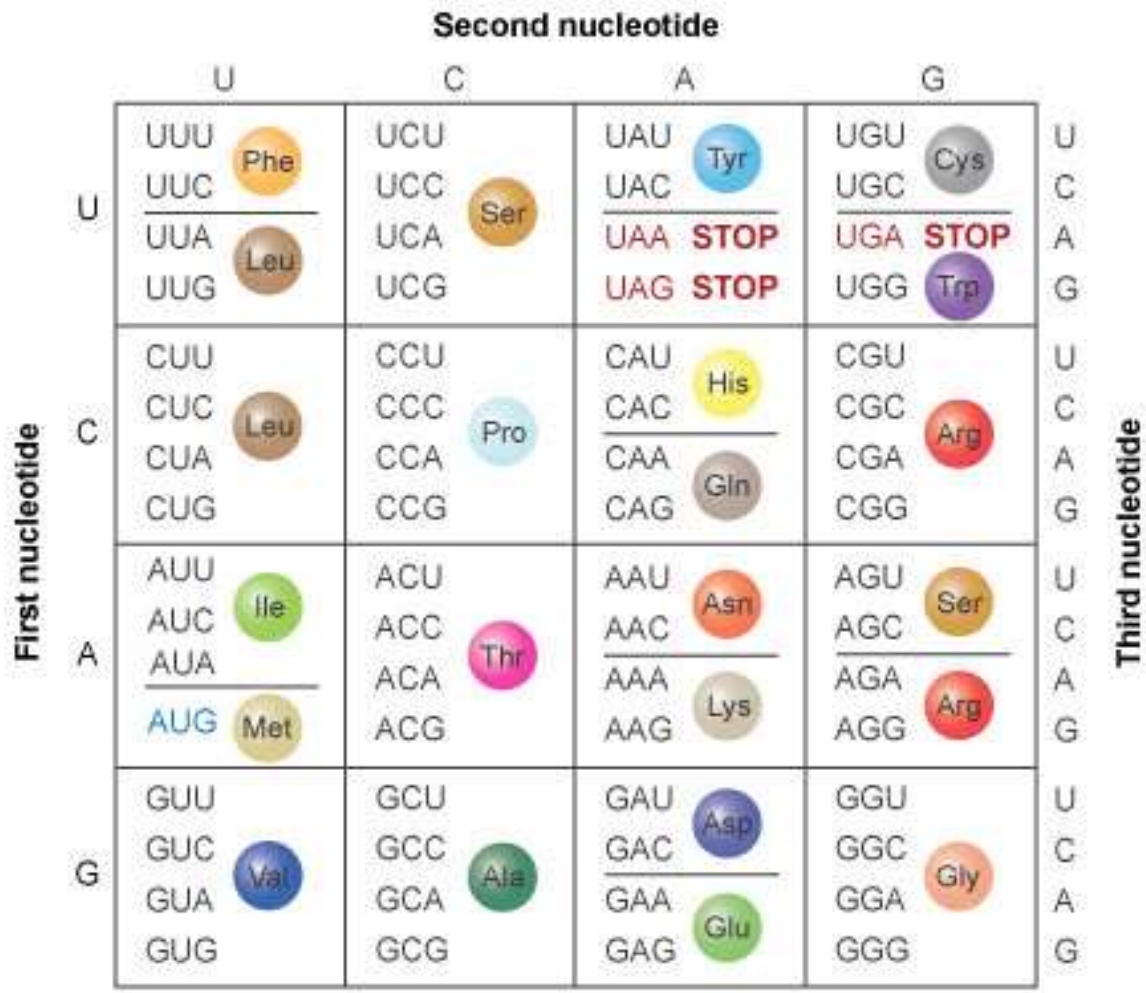


Possible reading frames:

AGG·TGA·CAC·CGC·AAG·CCT·TAT·ATT·AGC

A·**GGT·GAC·ACC·GCA·AGC·CTT·ATA·TTA**·GC

AG·**GTG·ACA·CCG·CAA·GCC·TTA·TAT·TAG**·C


Second nucleotide
U
C
A
G
First nucleotide
Third nucleotide
UUU Phe
UUC
UUA Leu
UUG
UCU
UCC Ser
UCA
UCG
UAU Tyr
UAC
UAA STOP
UAG STOP
UGU Cys
UGC
UGA STOP
UGG Trp
CUU
CUC Leu
CUA
CUG
CCU
CCC Pro
CCA
CCG
CAU His
CAC
CAA Gln
CAG
CGU
CGC Arg
CGA
CGG
AUU
AUC Ile
AUA
AUG Met
ACU
ACC Thr
ACA
ACG
AAU Asn
AAC
AAA Lys
AAG
AGU Ser
AGC
AGA Arg
AGG
GUU
GUC Val
GUA
GUG
GCU
GCC Ala
GCA
GCG
GAU Asp
GAC
GAA Glu
GAG
GGU
GGC Gly
GGA
GGG

What type of mutations can Effect fusion protein expression:

- Mutations that insert or delete individual bases cause a shift in the triplet sets after the site of mutation.
- Combinations of mutations that together insert or delete 3 bases (or multiples of three) insert or delete amino acids but do not change the reading of the triplets beyond the last site of mutation.

1. Point mutation

- The nature of the code predicts that two types of mutations will have different effects.
- If a particular sequence is read sequentially,

such as UUU AAA GGG CCC (codons)

a a l aa2 aa3 aa4 (amino acids) then a point mutation will affect only one amino acid.

For example, because only the second codon has been changed, the substitution of an A by some other base (X) causes aa2 to be replaced by aa5:

UUU AAX GGG CCC

aal aa5 aa3 aa4

2. Insertions and deletions:

a mutation that inserts or deletes a single base will change the triplet sets for the entire subsequent sequence.

A change of this sort is called a frameshift.

An insertion might take the following form:

UUU AAX AGG GCC C

aal aa5 aa6 aa7

Because the new sequence of triplets is completely different from the , the entire amino acid sequence of the protein is altered beyond the site of mutation.

Frameshift mutations are induced by the acridines, compounds that bind to DNA and distort the structure of the double helix, causing additional bases to be incorporated or omitted during replication.

Each mutagenic event sponsored by an acridine results in the additio

GET YOUR CLONED GENE SEQUENCED

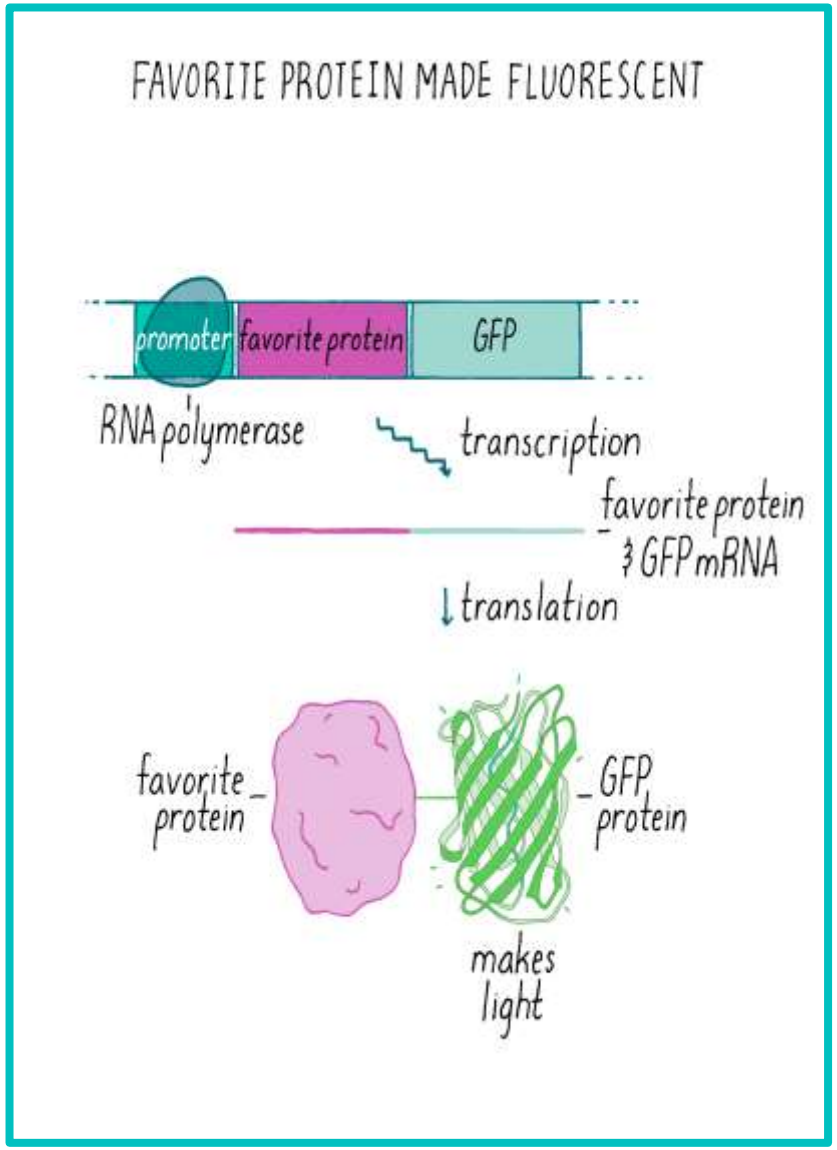
Frameshift mutatio

- The structure and/or enzymatic activity of each protein follows from its primary sequence of amino acids. By determining the sequence of amino acids in each protein, the gene is able to carry all the information needed to specify an active polypeptide chain.
- In addition to sequences that code for proteins, DNA also contains certain sequences whose function is to be recognized by regulator molecules, usually proteins. Here the function of the DNA is determined by its sequence directly, not via any intermediary code.
- Both types of regions, genes expressed as proteins and sequences recognized as such, constitute genetic information.
- In any given region, only one of the two strands of DNA codes for protein, so we write the genetic code as a sequence of bases (rather than base pairs).
- The identification of a lengthy open reading frame is taken to be prima facie evidence that the sequence is translated into protein in that frame.
- An open reading frame (ORF) for which no protein product has been identified is sometimes called an unidentified reading frame (URF)

Why do we need fusion proteins?

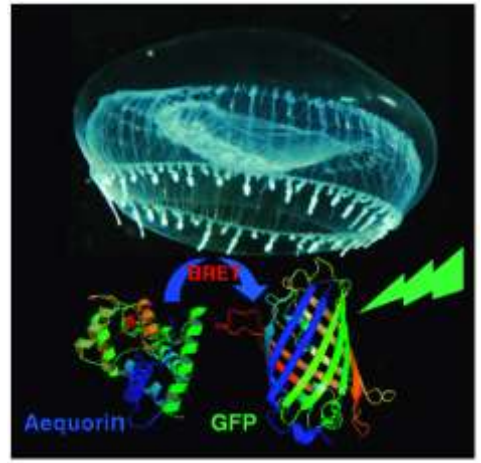
1. Two proteins need to be expressed together
2. Check expression of protein
3. Check localisation of protein in a cell
4. Chaperones co expressed with target protein

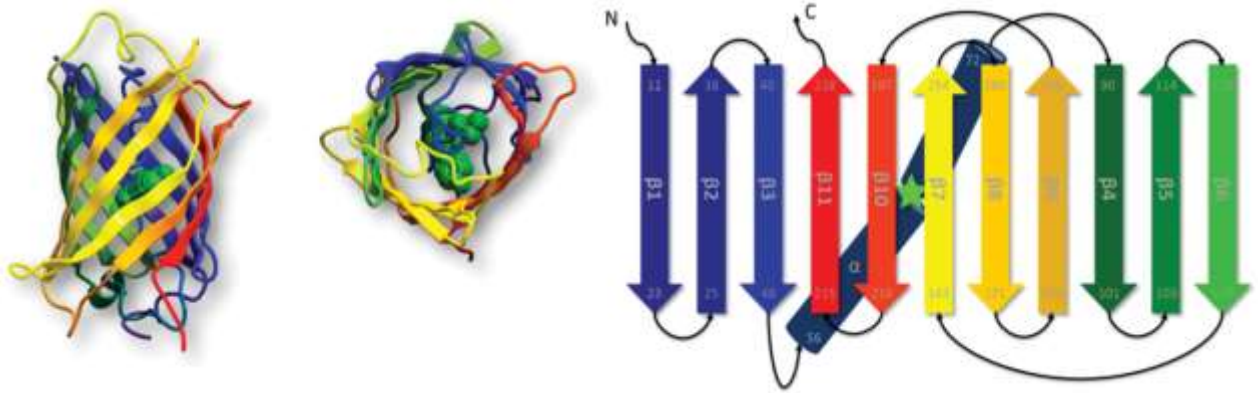
In-frame protein fusion and use of GFP



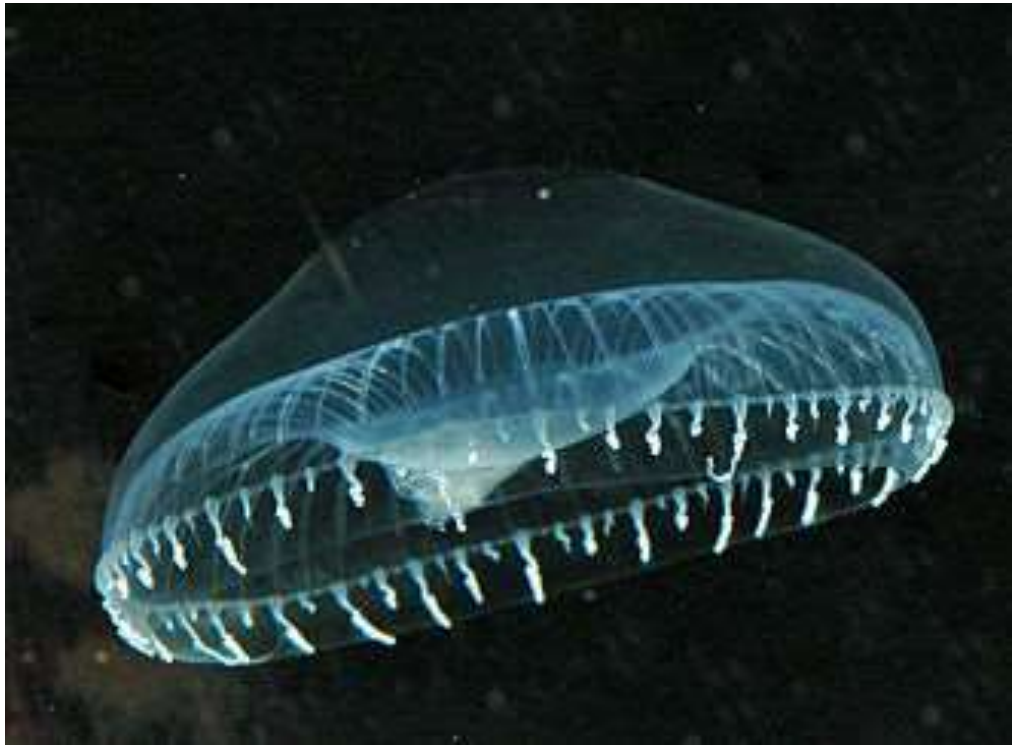
The first fluorescent protein to be cloned and characterized was green fluorescent protein (GFP) from the jellyfish species *Aequorea Victoria.*

It has **238 amino acid residues** and a green fluorophore which is comprised of only three amino acids: Ser65–Tyr66–Gly67.





Tertiary structure of GFP as determined by x-ray crystallography (PDB code 2B3P)



Aequorea victoria, also sometimes called the crystal jelly, is a bioluminescent hydrozoan jellyfish, or hydromedusa, that is found off the west coast of North America.

The species is best known as the source of two proteins involved in bioluminescence, aequorin, a photoprotein, and green fluorescent protein (GFP).

Why GFP?

1. Using the GFP as a marker gene.

2. Additions to either end of the GFP polypeptide are well tolerated and GFP fusion proteins still glow green under UV light.

3. It is detectable rapidly after the gene transfer without any prior staining procedures

4. Observation/sorting of GFP-expressing cells with fluorescence microscopy or flow cytometry leaves most cells unharmed, thus allowing almost continuous monitoring of the expression

5. GFP can be fused to other proteins to yield therapeutic fusion proteins, whose expression and subcellular location can be studied

Gene & Protein sequence of GFP

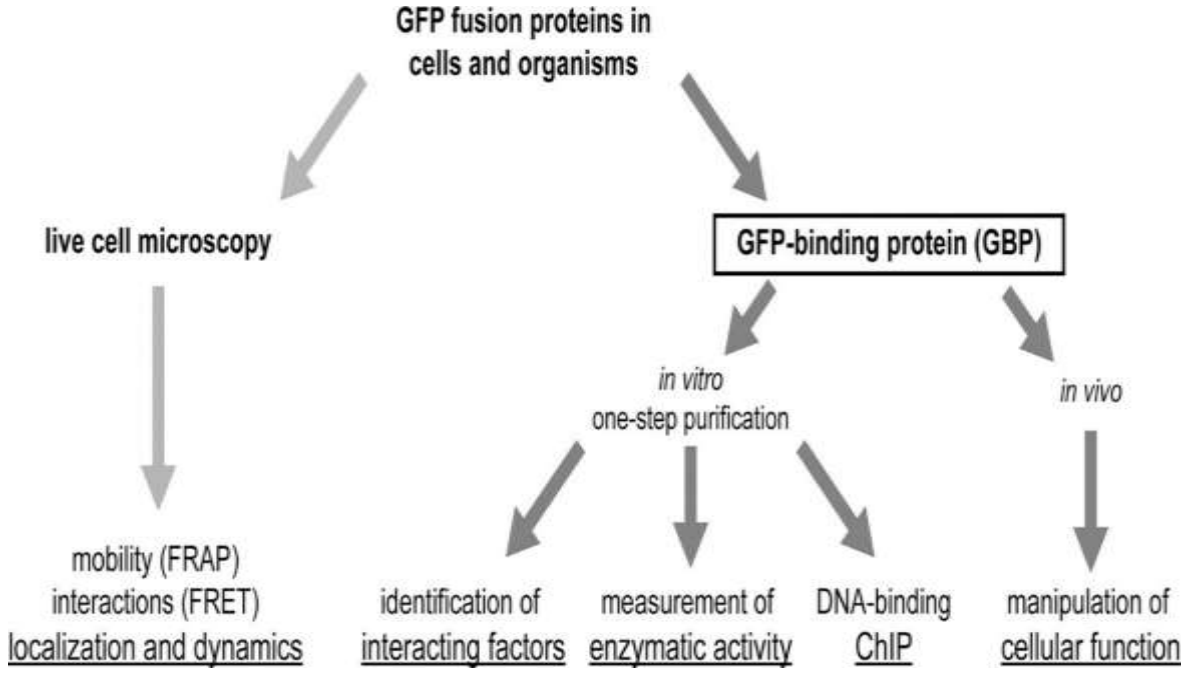
>ENA|AAA27722|AAA27722.1 Aequorea victoria green-fluorescent protein

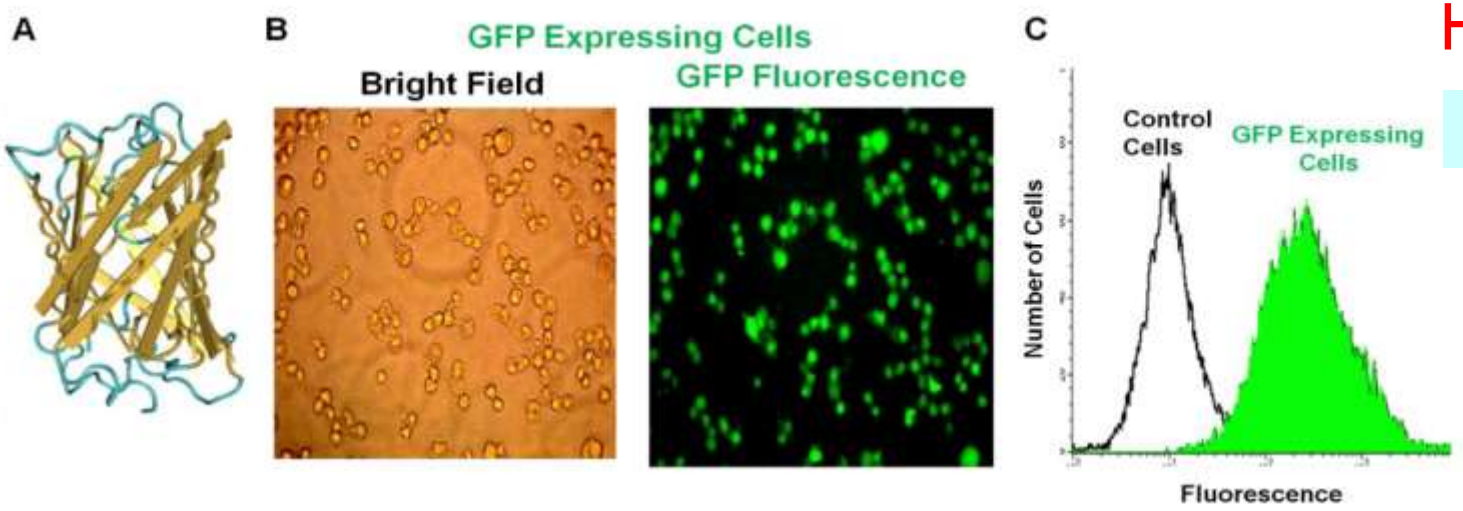
ATGAGTAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGT
GATGTTAATGGGCACAAATTCTCTGTCAGTGGAGAGGGTGAAGGTGATGCAACATACGGA
AAACTTACCCTTAAATTTATTTGCACTACTGGAAAGCTACCTGTTCCATGGCCAACACTT
GTCACTACTTTCTCTTATGGTGTTCAATGCTTTTCAAGATACCCAGATCATATGAAACAG
CATGACTTTTTCAAGAGTGCCATGCCCGAAGGTTATGTACAGGAAAGAACTATATTTTAC
AAAGATGACGGGAACTACAAATCACGTGCTGAAGTCAAGTTTGAAGGTGATACCCTCGTT
AATAGAATTGAGTTAAAAGGTATTGATTTTAAAGAAGATGGAAACATTCTTGGACACAAA
ATGGAATACAACTATAACTCACACAATGTATACATCATGGCAGACAAACAAAAGAATGGA
ATCAAAGTTAACTTCAAAATTAGACACAACATTGAAGATGGAAGCGTTCAACTAGCAGAC
CATTATCAACAAAATACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGACAACCATTAC
CTGTCCACACAATCTGCCCTTTCCAAAGATCCCAACGAAAAGAGAGATCACATGATCCTT
CTTGAGTTTGTAACAGCTGCTGGGATTACACATGGCATGGATGAACTATACAAATAA

>sp|P42212|GFP_AEQVI Green fluorescent protein OS=Aequorea victoria OX=6100 GN=GFP PE=1 SV=1

MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTL
VTTFSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV
NRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLAD
HYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITHGMDELYK

Application of GFP fusion proteins





How to monitor GFP?

Chimeric Construct with green fluorescent protein (GFP

GFP represents a good candidate for reporter gene construct. (A) 3-D structure of green fluorescent protein, (B) COS-7 Cells expressing GFP and (C) Flow cytometric analysis of GFP expressing COS-7 Cells

DESIGN OF A FLUORESCENT FUSION PROTEIN

Fluorescent protein	Excitation maximum (nm)	Emission maximum (nm)	Relative brightness
wtGFP	397	508	20,448 (0.6×)
EGFP	490	509	33,000 (1×)
PAGFP (preactivation)	504	515	2691 (0.08×)
PAGFP (postactivation)	400	517	13,746 (0.4×)
EYFP	514	527	51,240 (1.5×)
Venus	515	528	52,554 (1.6×)
Citrine	516	529	58,520 (1.8×)
ECFP	433	476	10,730 (0.3×)
Cerulean Blue	433	475	26,660 (0.8×)
DsRed	558	583	45,030 (1.4×)
T1	554	586	12,642 (0.4×)
mRFP	584	607	11,000 (0.3×)
Kaede (preactivation)	508	518	78,400 (2.4×)
Kaede (postactivation)	572	582	19,932 (0.6×)
HcRed	592	645	na
KFP-1 (preactivation)	na	600	<123 (0.004×)
KFP-1 (postactivation)	580	600	4130 (0.1×)
mAzami-Green	492	505	33,858 (1×)

To improve the utility of GFP, the codon bias was modified to make the protein more suitable for expression in mammalian cells, and two amino acids were mutated to enhance brightness and protein folding (S65T and P64L). The resulting variant is termed **EGFP** and is available from Clontech.

Point mutations can alter the spectral properties of EGFP to create blue (ECFP) (C for Cyan) and yellow (EYFP) (Y for yellow) variants, also available from Clontech. ECFP is substantially dimmer than EGFP, while EYFP is slightly brighter. To calculate the brightness of an FP, multiply the protein's extinction coefficient (ε, the efficiency of photon absorption) by its quantum efficiency (ϕ, the ratio of photons emitted to photons absorbed)

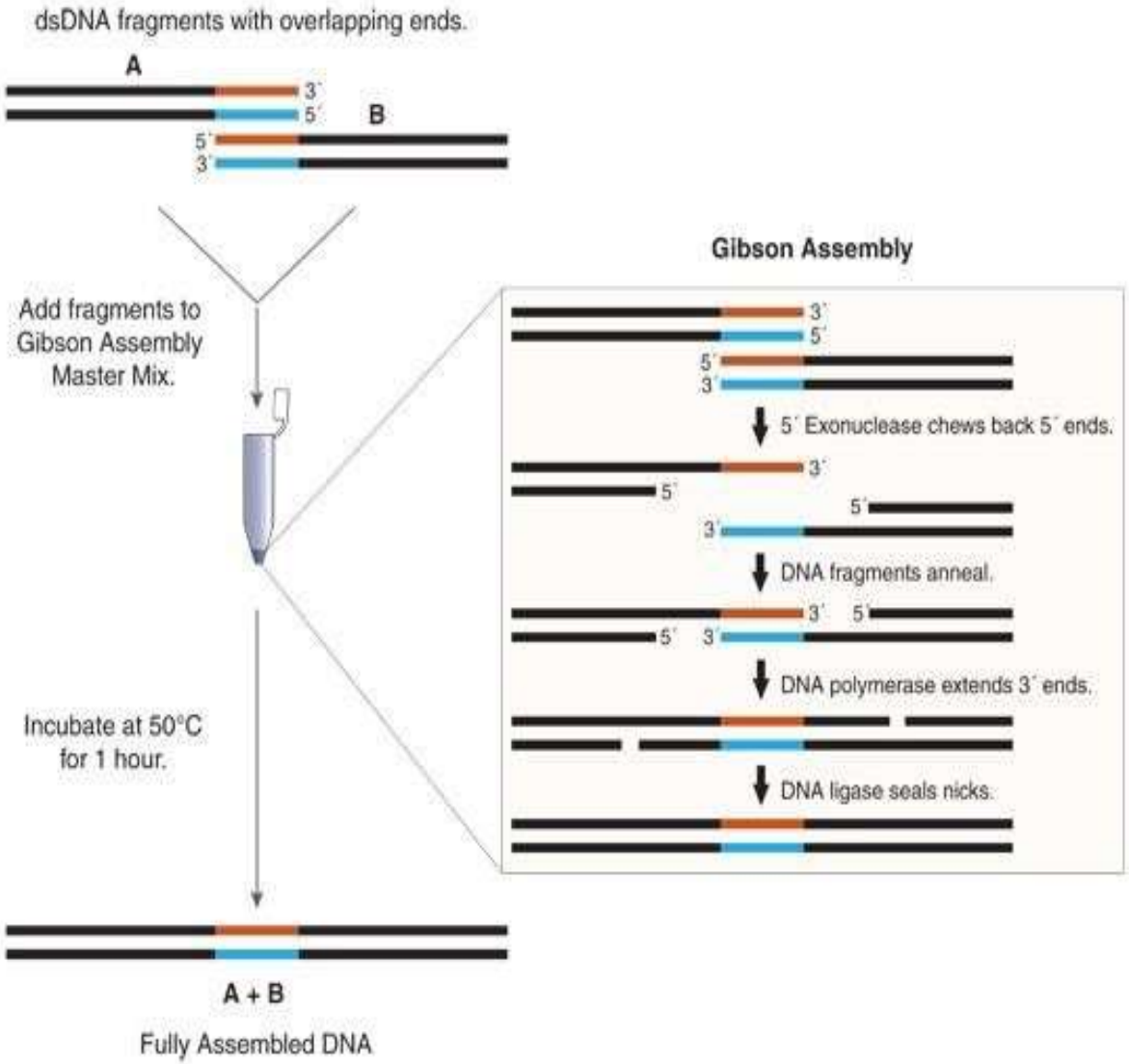
Assembly-PCR

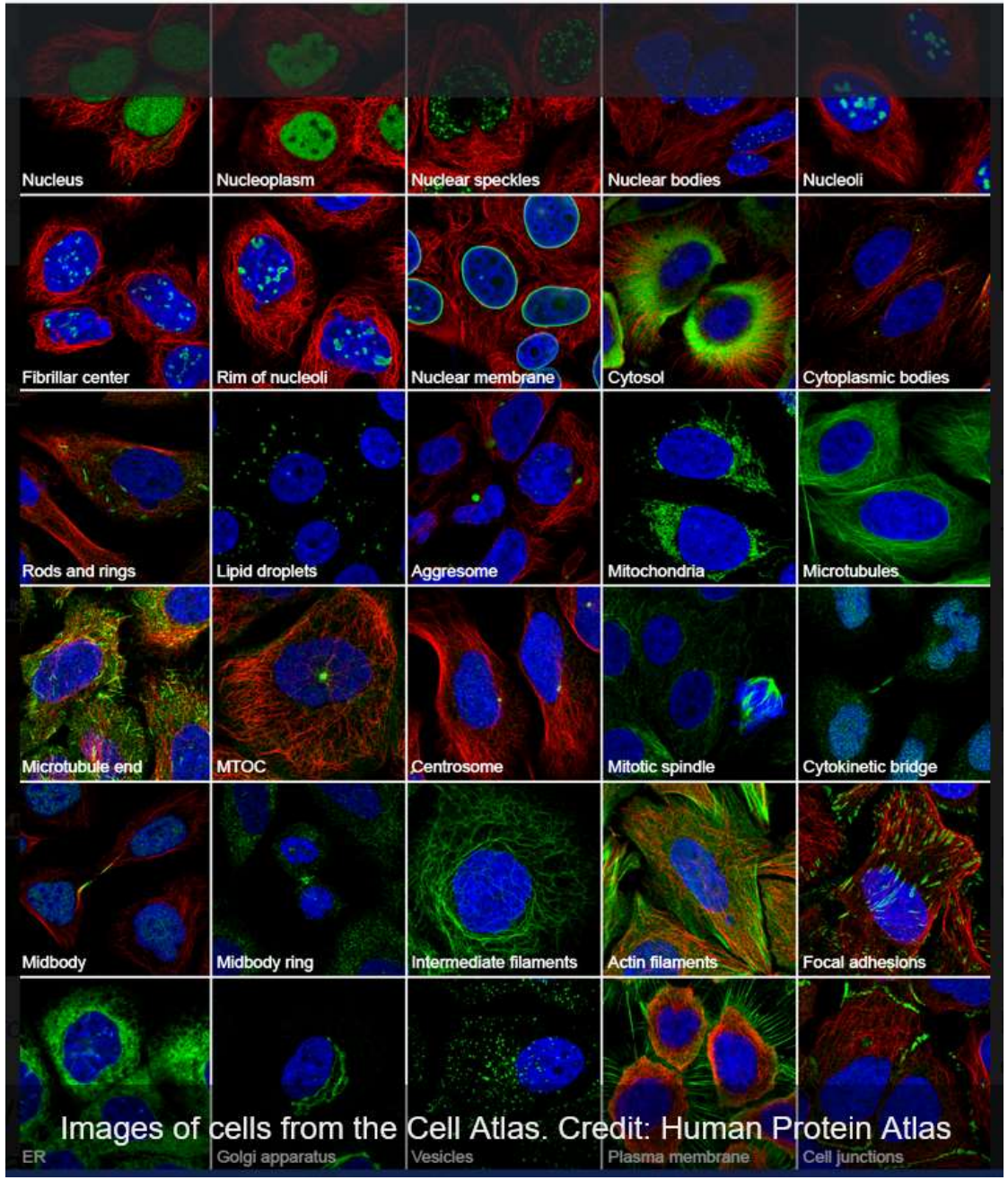
(*Polymerase Cycling Assembly* or *PCA*)

1.In this type synthesis of long DNA structures by performing PCR on a pool of long oligonucleotides with short overlapping segments, to assemble two or more pieces of DNA into onepiece.

2. Both the gene sequences should be known

3.It involves an initial PCR with primers that have an overlap and a second PCRusing the products as the template that generates the final full-length product.





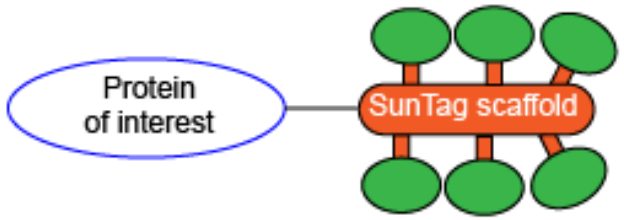
Images of cells from the Cell Atlas. Credit: Human Protein Atlas

Protein subcellular localization prediction involves the prediction of where a protein resides in a cell, its subcellular localization visualization through antibodies!

Traditional FP fusion protein

Low signal:
Only one copy of GFP is directly fused to the protein of interest

SunTag fusion protein scFV-GFP



High signal:
The SunTag scaffold recruits up to 24 copies of scFV GFP

Comparing traditional GFP fusion proteins to SunTag fusion proteins. A traditional GFP fusion fuses one copy of GFP to a protein of interest. Rather than fuse GFP to a protein, SunTag fusions contain a synthetic scaffold that recruits GFP fused to the scFV antibody.

Which of the following is important for the design and use of a reporter GFP fusion protein plasmid to study the subcellular localization of a protein of interest? Select all the apply.

- [ ] The GFP gene should be downstream of the promoter of the gene of interest.
- [ ] The GFP gene should be in the same reading frame as the gene of interest.
- [ ] The construct with the GFP gene and the gene of the protein of interest must be transduced into cells.
- [ ] the STOP codon from the gene of the protein of interest must be removed before insertion into the MCS upstream of the GFP gene.