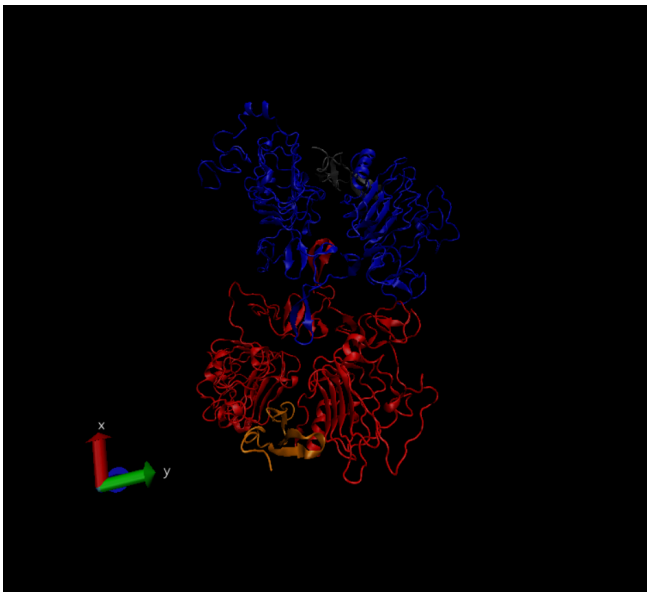
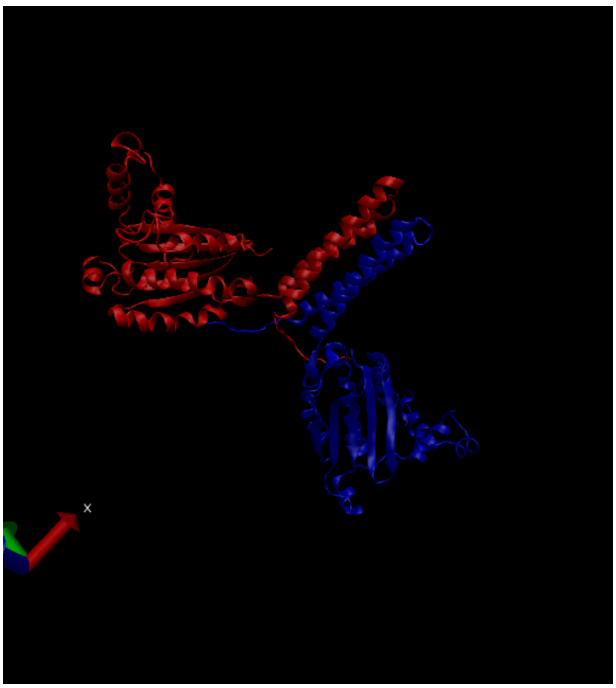
Task 1



EGFR Molecule in extracellular domain

TransMembrane Domain



Asymmetric dimer of kinase domains

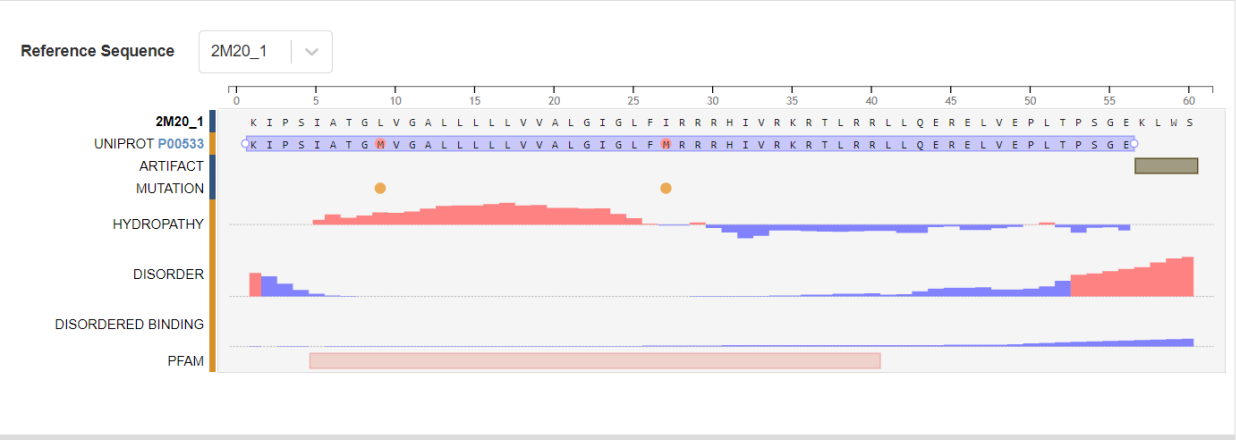
ERFG is usually overexpressed or mutated in different types of cancers. This leads to uncontrolled tumor growth. By targeting the EGFR with different inhibitors that can block its signaling pathways, we can prevent the spread of these tumors which could be used as a possible therapeutic strategy.

In the extracellular domain there are four key chains. In chains one and three there are leucine rich regions that are directly bound to other ligands like the epidermal growth factor and the transforming growth factor. In chains two and four, these chains are rich in cysteine which helps stabilize the structure of the receptor,and it helps facilitate dimerization upon ligand binding. As ligands bind it results in a conformational change which results in the receptors dimerization which is crucial for the receptor to activate properly.

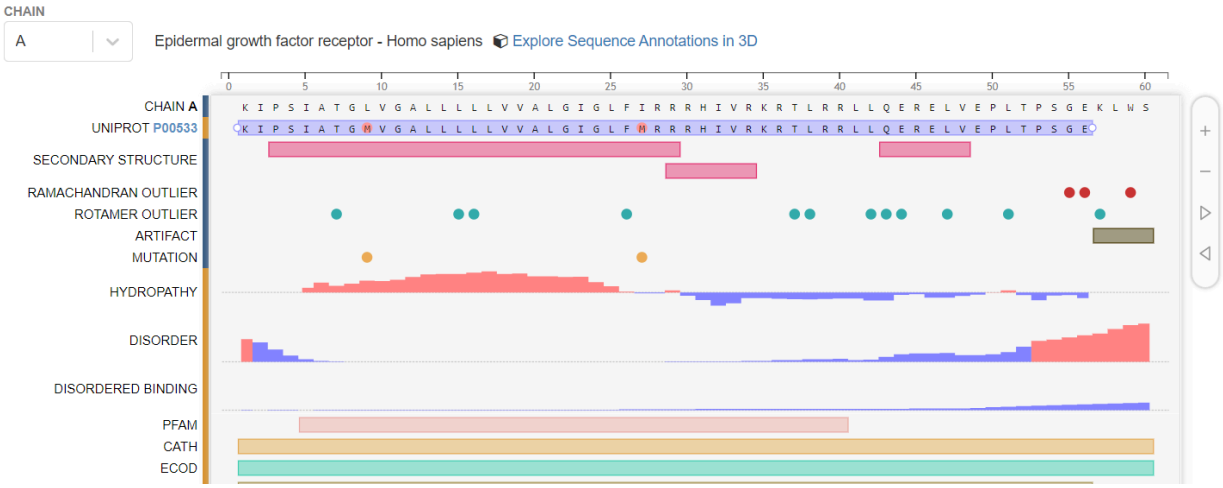
The transmembrane domain is made up of two different dimers that connect with one another. This creates a connection between the extracellular and intracellular parts of the molecule.

The intercellular domain contains our asymmetric dimer of the kinase domains. This domain is where the receptor's enzymatic activity occurs. Upon dimerization it activates many downstream signaling pathways like MAPK, AKT, and JAK/STAT pathways. This region has many tyrosine residues that serve as a docking sight for signaling proteins when phosphorylated,. [1]

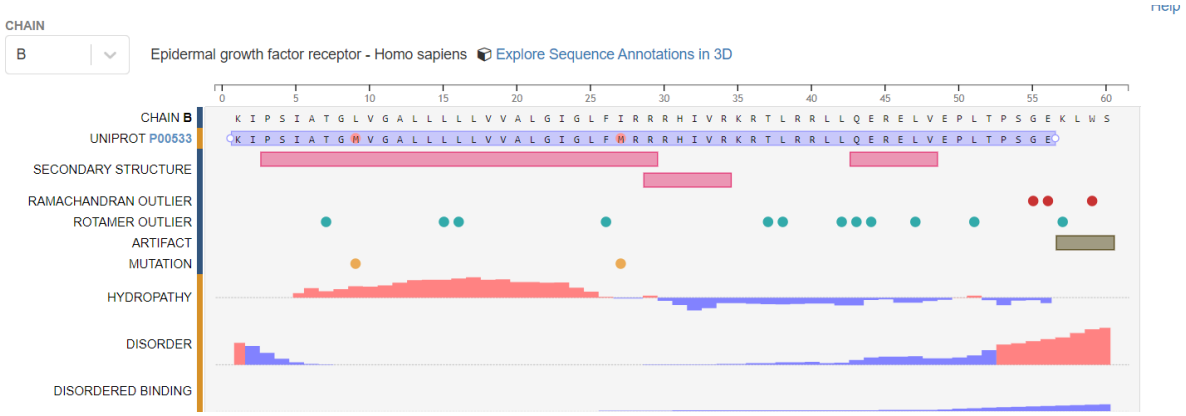
Hydropathy Plot for the transmembrane domain:

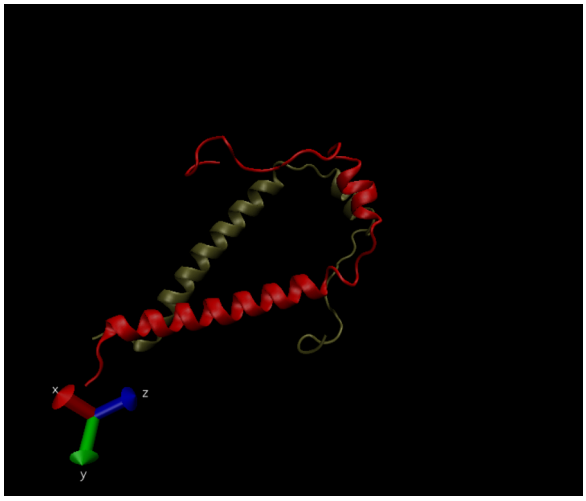


Hydropathy plot chain A:



Hydropathy plot chain B:



Structure: 

The transmembrane segment is typically a hydrophobic alpha helix which is what allows it to embed itself into the cell membrane. When the ligand binds to the extracellular domain it results in the dimerization of the transmembrane domain, This is what activates the intracellular kinase domains,

According to the overall hydrophobicity chart of the transmembrane domain we can see that it has a large hydrophobicity from residues 1 to 25. This could be because of the high concentration of leucine and valine amino acids between these residues. These amino acid groups have very hydrophobic properties as they are nonpolar.

As explained previously ERFG is usually overexpressed or mutated in different types of cancers. This leads to uncontrolled tumor growth. Ligands bind to the extracellular domain of the EGFR and it induces changes that result in dimerization which leads to activation. By targeting the EGFR with different inhibitors that can block its signaling pathways, we can prevent the spread of these tumors which could be used as a possible therapeutic strategy.

Task 2
(A) Analyze the EGFR ligand (EGF) binding site

Levels of organization

Primary Structure (1˚)
ECPLSHDGYCLHDGVCMYIEALDKYACNCVVGYIGERCQYRDLKWWE

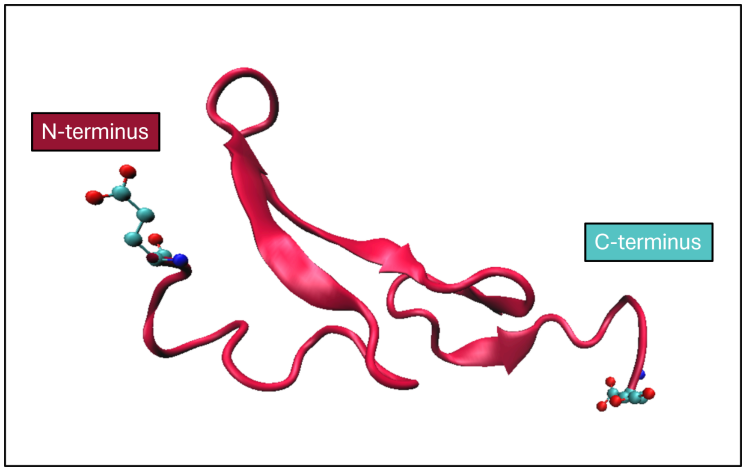
Backbone contains the following amounts from each classified groups of amino acids on the basis of charge:
Positively charged: Lys, Arg and His. (15.1 %)
Negatively charged: Asp and Glu. (17.0 %)
Uncharged polar: Ser, Thr, Asn, Gln. (7.5 %)
Hydrophobic side chains: Ala, Val, Ile, Leu, Met, Phe, Tyr, and Trp. (30.2 %)

From above the overall charge is balanced with about the same percentage between negative and positive side chains. There is a fair amount of the peptide that is hydrophobic indicating possible association with membrane or specific folding to minimize contact with water.

Secondary Structure (2˚)

Favorable Amino Acids in helix structure formation include Ala, Leu, Met, Gln, Ly, and Glu. Although these amino acids are 28% in abundance in the above protein sequence, they don't come sequentially. Additionally, the polar amount of amino acids is 7.5 % indicating less making of hydrogen bonds. Thus it is predictive of the primary structure percentage of hydrophobic side chains that beta sheets are the most likely to be formed which agrees with structure in image above.

Tertiary Structure (3˚)



Folded into a compact chain which for its signifies its hydrophobic interactions with the surrounding.

(B) Design a protein/peptide that is 15-25 residues long and can bind to the EGFR active site and inhibits its function. Explain your work.

When designing protein to bind to an active site, it may be considered either the amino acids which constitute the region where a ligand best bind and from there design a peptide which contains amino acids that complement that of the binding site. On the other hand, a ligand may be considered and thus the peptide made similar to the ligand so as to compete with it.

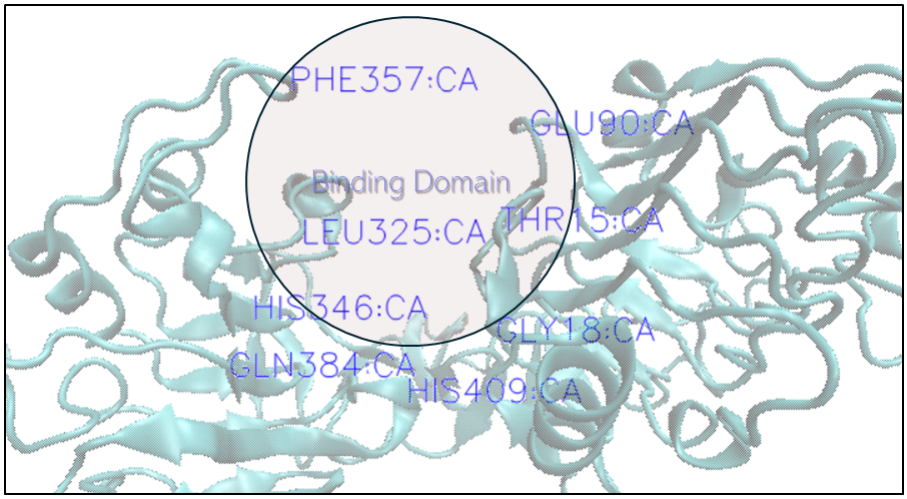
Binding domain most common amino acids include SER, GLN, and LEU. Those which are typically found at the end of an alpha helix, since the helix formed in structure given are not in maximum contact with the active site (their ends instead of their sides are in greater contact with ligand). Additionally as is the case with all helices formed, the atoms are occupied in hydrogen bonding within the helix.

Thus 1st method is choosing by complement favor the following amino acids,

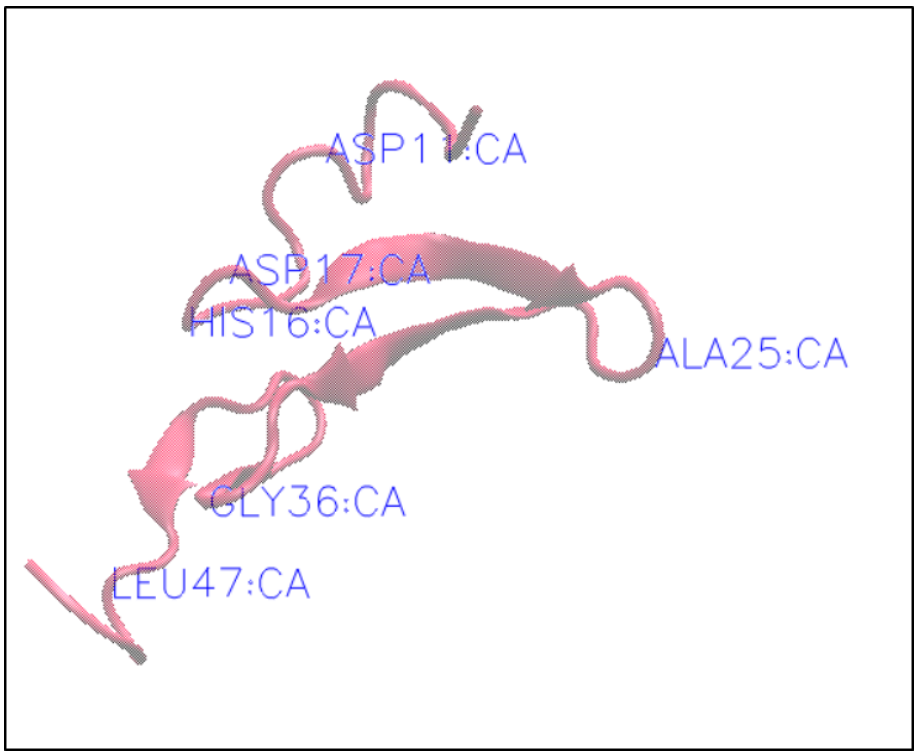
Serine (SER) Complements Threonine (THR)

Glutamine (GLN) Complements Asparagine (ASN)

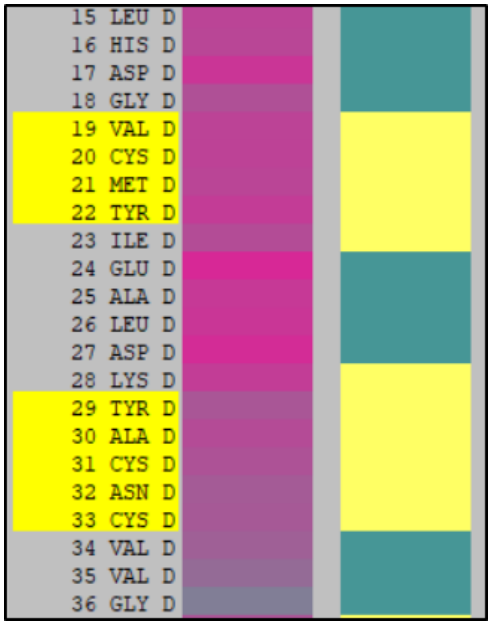
Leucine (LEU) Complement: Alanine (ALA)



Second method is to select interacting amino acids which most interact and increase their abundance in the constructed peptide. These include ASP, HIS, ALA, and GLY.



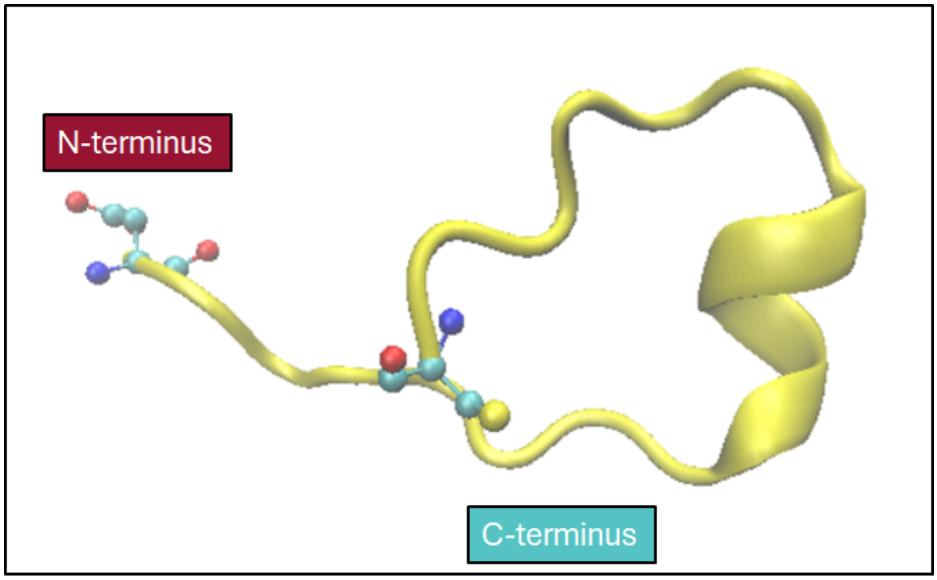
Additional consideration is the making of beta strands which contribute the ligand fitting into the pocket (active site). The sequence of above beta strands



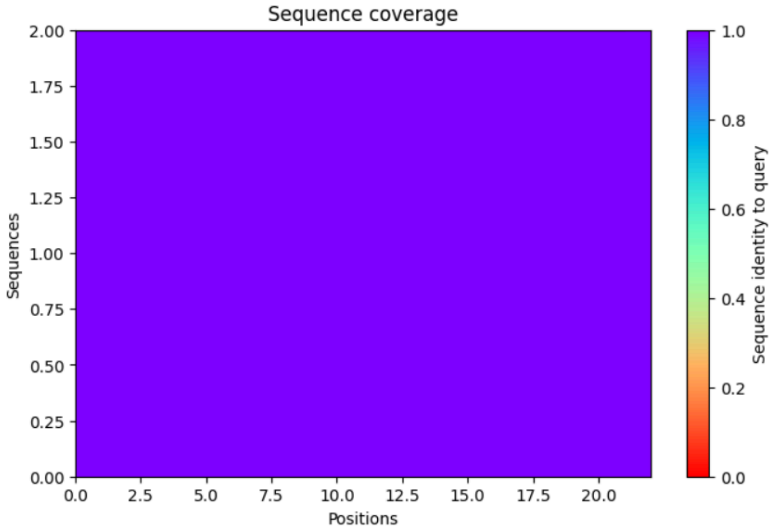
Thus for the purpose of making competing peptides, few amino acids are chosen which contribute to the making of beta strands and in between them and at their ends, amino acids which complement the active site. Thus,

ASP ASN THR ALA **VAL CYS MET** THR ASP **TYR ALA CYS ASN** LEU GLY THR ASN GLY CYS ASN CYS MET

(C) Predict the structure of your inhibitor protein/peptide using AlphaFold and include it in your report.



The above structure prediction was made via an optimized version of AlphaFold called ColabFold [AlphaFold2.ipynb - Colab]. Since alphafold took much time and could not return results. From their feedback statistical model, every sequence was accounted for as plotted in the following,



Structure does not exactly resemble the intended one as there were not any beta strands formed. Although given it is shorted, it allows it to spread further and thus have a greater surface to volume ratio and may get into further regions where an otherwise bigger peptide is hindered by its side chains.

(D) After designing your inhibitor protein/peptide, you will express it in bacteria using circular plasmids for manufacturing.

Peptide sequence:
ASP ASN THR ALA **VAL CYS MET** THR ASP **TYR ALA CYS ASN** LEU GLY THR ASN GLY CYS ASN CYS MET (DNTAVCMTDYACNLGTNCNC)

Due to the redundancy in genetic code (several codons code for the same amino acid), thus there are multiple RNA sequences corresponding to the above peptide. One possible corresponding RNA sequence:
GAU AAU ACU GCU GUU UGU AUG ACU GAU UAU GCU UGU AAU UUA GGC ACU AAU GGC UGU AAU UGU AUG

Corresponding DNA sequence:
5’GATAATACTGCTGTTTGTATGACTGATTATGCTTGTAATTTAGGCACTAATGGCTGTA ATTGTATG 3’

Forward primer
5’ GATAATACTGCTGTT 3’

Could even be shorter given that the DNA sequence is very short so with the above primer, specificity is high. Primer melting temperature is 36.7 centigrade. It is 15 nucleotides long. GC content 33.33% which is not optimal though the entire DNA sequence does not contain more than 40% so this is a limitation specific to the DNA sequence and not the primer.

Reverse primer
5’ CATACAATTACAGCC 3’

Primer melting temperature is 38.8 centigrade. It is 15 nucleotides long. GC content 40%.

PCR Protocol

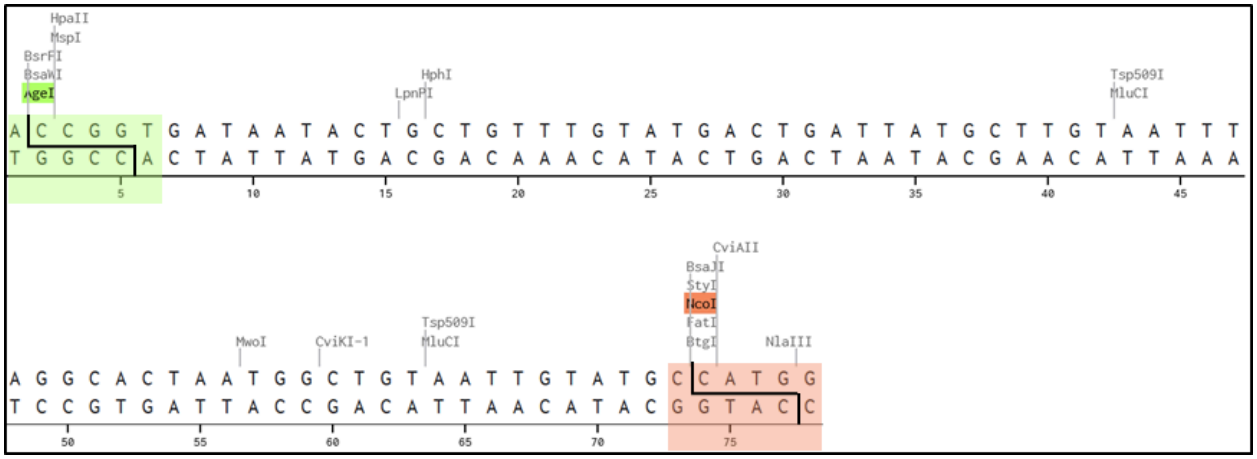
Materials:
Template DNA, Forward primer, Reverse primer, Taq DNA polymerase, dNTPs, PCR buffer and Sterile water if diluting is necessary.

Repeating steps past denaturation for 3-5 min:

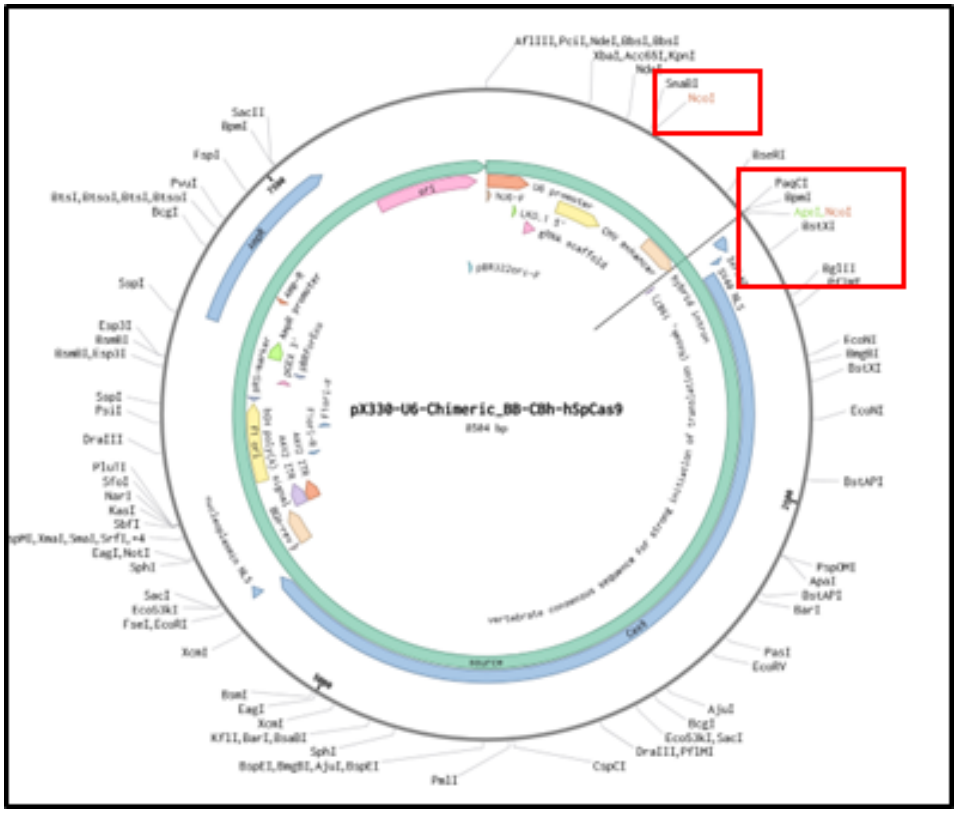
- Denaturation: ~90-99°C for 30 seconds
- Annealing: 30°C for around a minute (Specific to forward and reverse primer melting point in that it has to be lower).
- Extension: Increase temperature slightly above 30 centigrade and leave for another minute

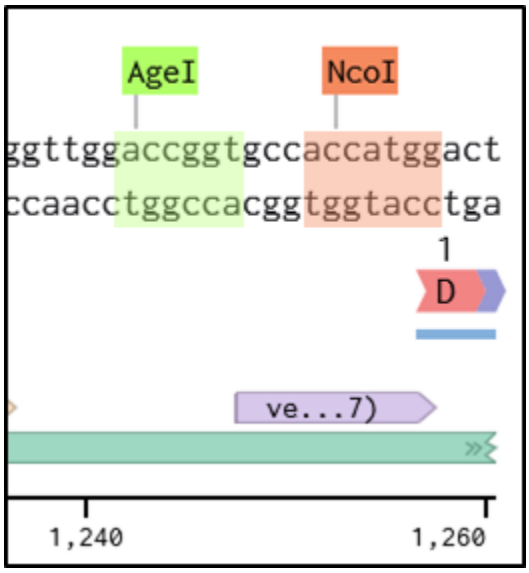
Plasmid used: **pUC19-SpCas9-2A-eGFP** [Plasmid]

Restriction sites
First several nucleotides on each end of each strand are so to complement AgeI and NcoI restriction enzymes also found within the plasmid. The highlighted areas in green and red are the new nucleotides added and the restriction enzymes are also shown. Both yield sticky each on both sides of the DNA.



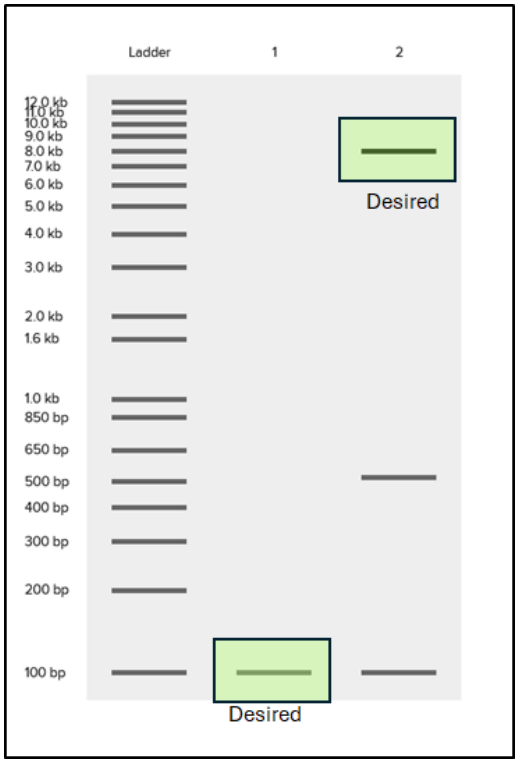
Correspondingly within the plasmid,





Although within the plasmid there are two sites where NcoI can cut, the following restriction site is available for AgeI which would in turn at times allow for the inhibitor gene to insert itself if the intended region between AgeI and NcoI would not happen all the time.

The following step is to gel purify the digested fragments, that is on the basis of the length of each fragment separate through say electrophoresis, the desired and digested inhibitor gene and the digested plasmid from everything else.



In column 1 in the above image, it represents the fragments made from the inhibitor DNA, while 2 is the plasmid. In column one, there should be two bars instead of one. But because the entire inhibitor and gene and added ends amount to 78 nucleotide pairs so it could not be differentiated from the above simulation. It should also prove difficult in actual experiments to differentiate between the fragments formed by this gene.

The proceeding step is to combine the two fragments of interest. That is to combine the linearized plasmid and the purified inhibitor gene in a ligation reaction via the use of some ligase enzyme. It comes past this step, the need to introduce combined plasmid and gene of interest into a host cell like E. Coli via heat shock to make the cellular membrane permeable. Next is to verify whether or not the gene of interest gets expressed.

Task 3

A) First it is determined the size of the inhibitor protein. Our chosen one consists of 20 residues. The average length of one extended residence is 0.38 nm. Thus the length of our peptide were it fully extended (maximum possible length) is 7.6 nm. Thus a small unilamellar vesicle (20-100 nm) may be used. In our chosen protein there are on average more hydrophilic amino acids than hydrophobic thus the drug will end up with LNP, that is the aqueous core. Several other things may be added to the surface of LNP such as polyethylene glycol to increase its lifetime as the immune mechanism in the body tries to counter it.

B) A synthetic peptide ligand called GE11 is made expressly to attach to the extracellular domain of EGFR. Its sequence is identified as YHWYGYTPQNVI. As it is known that EGFR stimulates growth, yield of cellular growth decreases when EGFR transmembrane - juxtamembrane (TM-JM) is linked to a high affinity ligand such as GE11. This can be incorporated into LNP which in addition has high affinity to interact with phospholipid membranes, it is more effective than other drug delivery NP when it comes to transmembrane proteins.

C) The best loading method may be hydrophobic core encapsulation. This works by incorporating the hydrophobic protein inhibitor into the lipid bilayer of the LNP. This method uses the interactions between the hydrophobic protein inhibitor and the lipid molecules to ensure encapsulation and protection from an environment that is aqueous.

Step by step proposed procedure:

1. Dissolve hydrophobic protein inhibitor and lipids in ethanol or chloroform (Any organic substance)
2. Mix to ensure an even distribution of inhibitor and lipids
3. Evaporate the solvent to form a thin lipid film
4. Hydrate the film using an aqueous buffer while vortexing to form the LNP's
5. Then suspend through polycarbonate membranes to achieve the targeted NP size

Sources

Huang, J., Moses, M. A., Bhuva, P., Plonowski, A., & Goldstein, A. M. (2021). A molecular mechanism for the generation of ligand-dependent differential outputs by the epidermal growth factor receptor. *eLife*, 10, e73218. https://doi.org/10.7554/eLife.73218

Nature Structural Biology 10, 980 (2003) doi: 10.1038/nsb1203-980

Acta Crystallographica Section D: Biological Crystallography 57, 389 (2001) doi: 10.1107/S090744490001898X

Nature 381, 456 (1996) doi: 10.1038/381456a0

AlphaFold2.ipynb - Colab, https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb#scrollTo=mbaIO9pWjaN0

Plasmid sequence, https://www.addgene.org/browse/sequence/419908/