In Vitro Construction of Epidermal Growth Factor Receptor Kinase Domain

A tool for selection and screening of tyrosine kinase inhibitors to improve non-small cell lung cancer treatment

Abstract:

Activating mutations of the epidermal growth factor receptor (EGFR) may lead to the development of non-small cell lung cancer (NSCLC). Response to inhibition by tyrosine kinase inhibitors (TKI's) is found to be highly various and depends on mutation phenotype. For many rare mutations, of which clinical data is lacking, optimal treatment is yet to be determined. E. coli systems hold potential for in vitro screening of TKI inhibition of rare EGFR mutations to improve NSCLC treatment, as these systems may offer a cheap and fast screening method. During this research project, an EGFR gene construct, encoding for residue 681 to 1211, was successfully cloned in a pETM-30 vector. After expression in E. coli strains optimized for recombinant protein synthesis, the receptor construct can be used for activity screening.

Rijksuniversiteit Groningen, Department of Drug Design (GRIP)
Bachelor's project

Author: Michiel Zietse (3001830)
Supervisors: Prof. Dr. Matthew Groves
Juliana Vilachá

January 23, 2020

Table of Contents

1. Introduction

In 2018, 1.76 million annual deaths were caused by lung cancer. Non-small cell lung cancer (NSCLC) accounts for approximately 80% of lung cancer cases. [1] Mutations of the epidermal growth factor receptor (EGFR) occur in 17% of these patients. [2] Activating EGFR-mutations may lead to the stabilization of the active conformation of EGFR, leading to continuously activation of various signaling pathways, as can be seen in figure 1. This may subsequently result in increasing cell growth, prolonged cell survival and the inhibition of pro-apoptotic protein synthesis. [3]

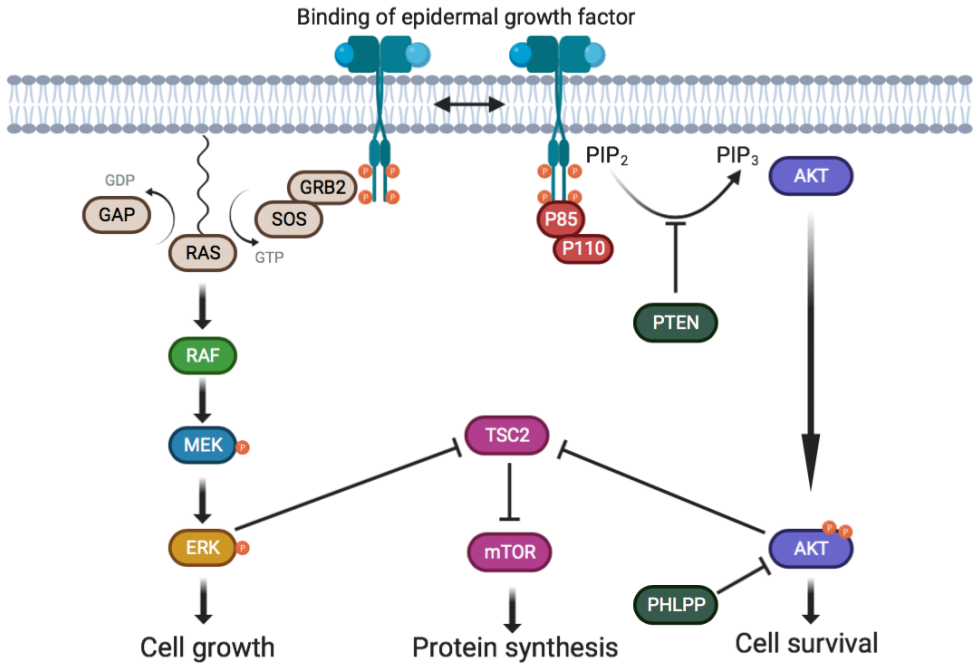

Figure 1: Signaling pathway of the epidermal growth factor receptor. Ligand binding of EGFR results in receptor dimerization, leading to phosphorylation of EGFR tyrosine residues. Conversion of GTP to GDP activates RAS, resulting in downstream kinase signaling via RAF/MEK/ERK and eventually cell growth. Also, the AKT-pathway may be activated, leading to increased cell survival. Moreover, activation of ERK and ATK may lead to inhibition of TSC2, canceling inhibition of anti-apoptotic protein synthesis via mTOR. The figure was created using Biorender.

EGFR belongs to the family of Receptor Tyrosine Kinases (RTKs). These receptors are composed of three domains: the extracellular-ligand binding domain, a transmembrane α-helix and a kinase domain. [4] The intracellular kinase domain has a bilobal structure. The N-lobe is formed by 5-stranded β-sheets (β1-5) and an alpha-helix (αC). A hinge region connects the N-lobe to a larger C-lobe, which mostly consists of α-helices and four β-strands (β6-9). Both lobes have different functional structures. In the N-lobe, both the alpha helix (αC) and P-loop play a role in correctly positioning ATP to be used for catalysis. The catalytic properties of the kinase originate from the catalytic loop in the C-lobe. Furthermore, the activation loop (A-loop) modulates kinase activity by blocking the substrate from binding the C-lobe. [5–7]

The most common EGFR mutations are exon 19 deletions (del19), exon 21 L858R, accounting for 45% and 40% of EGFR mutations respectively. [8] In L858R, a single point mutation leads to the transcription of arginine instead of leucine in exon 21. This mutation occurs in the A-loop, leading

to a shift in the protein conformational equilibrium resulting in a constitutively active receptor. [9] In del19, the αC-helix is shortened by 3 to 8 residues, also resulting in increased kinase activity.[8]

Three generations of tyrosine kinase inhibitors (TKIs) have been developed, of which five individual TKIs have been approved by the EMA. Currently, gefitinib, erlotinib (first-generation) and afatinib (second-generation) are being used as a first-line treatment in advanced NSCLC. These TKIs target the EGFR ATP-pocket, where first generation TKIs show reversible binding. In both del19 and L858R, receptor ATP-affinity is reduced. Therefore, first-generation TKIs can selectively bind mutated EGFR, sparing wild-type (WT) EGFR. [10] Second generation afatinib occupies the ATP-pocket by covalently binding Cys^{797}, therefore blocking kinase activity. [11,12]

After a period of progression-free survival (PFS), resistance to first- and second-generation TKIs is commonly observed. Substitution of threonine for methionine in position 790 (T790M), also referred to as a gatekeeper mutation, accounts for 50% of resistance cases. [11] Resistance may be the result of an increase in ATP-affinity in these secondary mutations. [13] Osimertinib, as a third-generation TKI, has proven efficacy in T790M-mutations. Just as afatinib, osimertinib covalently binds the Cys^{797}-residue. Furthermore, osimertinib has also shown to be effective in first-line treatment of L858R and del19. [14] However, resistance to osimertinib treatment has been observed by substitution of residue 797 cysteine by serine (C797S).

According to the January 2019 EMSO clinical practice guidelines, either erlotinib, gefitinib or afatinib can be used as a first-line treatment. [15] Nonetheless, variable responses for various EGFR-mutation subtypes have been observed. [8] Precise characterization of tumor genomics has been enabled by technological advances, such as next-generation sequencing (NGS). [16] These techniques allow for a more personalized approach to TKI selection in order to reach optimal treatment.

In table 1, the IC50's (nM) of in vitro studies for inhibition of various EGFR mutations by TKI's in Ba/F3 cells were summarized by *Kobayashi et al. (2016).* It can be seen that inhibition of EGFR by TKI is EGFR-mutation specific. For example, In del19 different deletional locations respond differently to inhibition by various TKIs. [8] Especially in more rare mutations, such as insertions in exon 20 (ins20) and exon 18 deletions (del18), response to treatment is highly variable and depends on the mutational subtype. [17–19] Ins20 are considered TKI resistant tumors. However, in vitro studies found that up to 20% of the ins20 subtypes can be successfully inhibited by first-generation TKIs. [18]

In vitro models of TKI pharmacodynamics were found to match clinical observations. [20] Therefore, in vitro models hold potential as a tool for screening of TKIs in different EGFR mutations. These screening results can then be used as a rationale for new clinical studies. [8,21] For example, an in vitro study by *Chiba, M. et al. (2017)* found that rare point-mutations L474S, D761Y and T854A in Ba/F3 cells were insensitive to first-line TKIs. It was found, however, that these seemingly resistant mutations could be inhibited by osimertinib. [22,23]

Table 1: Summary of in vitro screening for inhibition of various EGFR mutations in mouse Ba/F3 cells by different TKI's. IC50 values (nM) are shown for TKI inhibition. TKI binding of <10 in blue, 10-99 in light blue, 100-999 in yellow and >1000 shown in red. Data was collected and table was published by by Kobayashi et al. (2016). [8]

Exon	Category	Mutations	First generation		Second generation			Third generation	
			Gefitinib	Erlotinib	Afatinib	Dacomitinib	Neratinib	Osimertinib	Rociletinib
18	Del18	delE709_T710insD	882	884	1.7	29	27	93	999
	E709X	E709K	187	215	0.7	16	6	62	706
	G719X	G719A	213	167	0.9	6	1.1	53	214
19	Del19	delE746_A750	4.8	4.9	0.9	<1	60	1.1	19
	Del19	delE746_S752insV	306	14	0.2	1.4	86		
	Del19	delL747_A750insP	7.4	13	1	1.6	30		
	Del19	delL747_P753insS	4.1	5.4	2	1.9	38		
	Del19	delS752_I759	35	7.9	0.2	2	6.7		
	Ins19	I744_K745insKIPVAI	400		7				
	Ins19	K745_E746insTPVAIK	100		0.9				
20	Ins20	A763_Y764insFQEA	174	48	3.7			44	673
	Ins20	Y764_V765insHH	>1000	3845	79			237	1730
	Ins20	M766_A767insAI		3403	79				
	Ins20	V769_D770insASV	3100	4400	72	230	48	333	5290
	Ins20	D770_N771insNPG	3356	3700	72		230	42	262
	Ins20	D770_N771insSVD		3187	86				
	Ins20	H773_V774insH		>10 000	268		550		
	S768I	S768I	315	250	0.7			49	
	T790M	T790M+delE746_A750	8300	>10 000	64	140		3	28
	T790M	T790M+L858R	>10 000	>10 000	119	300		21	13
21	L858R	L858R	26	16	4	2.6	1.4	9	140
	L861Q	L861Q	170	103	0.5		3.3	9	

In table 1, it can be observed that there are still many blanks, meaning in vitro screening has yet to be completed for many combinations of mutations and TKI inhibition. In fact, over 2794 unique EGFR mutations have been clinically observed, according to the COSMIC database. [24] Therefore, there are still many mutations waiting to be screened for inhibition by various TKI's, in order to find the optimal treatment. A disadvantage to the summary of results provided in table 1, is the fact that inter-laboratory variation and methodological variation should be considered. As data are generated by different studies, inhibition of various mutations cannot be compared directly. This variation should especially be considered for rare mutations, as there are a limited number of studies.

A better comparison can be obtained by In vitro screening of multiple mutations following the exact methodology. In the case of rare mutation screening, a few common mutations of which in vitro data is abundant, should also be included in the study as a comparison. However, this will lead to an increase in workload. Therefore, a screening method with an increased throughput is needed. *Elloumi-Msedi et. al. (2013)* have previously succeeded in expressing the active recombinant EGFR intracellular domain in Escherichia coli (E. coli). Dimerization of the receptor, on which kinase activity depends, was accomplished by fusion to glutathione S-transferase (GST). Based on tyrosine kinase activity assays using [γ-^{32}P]ATP, the EGFR dimer construct proved sensitive to inhibition by genistein. [25,26] These systems could, therefore, potentially be used for in vitro screening of TKI inhibition of rare EGFR mutations, of which clinical data is currently lacking. [26] E. coli-based protein expression may offer a cheap and fast screening method.

During this research project, the main goal is to successfully synthesize Wild-Type EGFR-kinase domain in E. coli.

2. Methods

2.1 Construction of plasmid

A pJP1520 vector containing the full EGFR gene (Homo sapiens) was purchased from DNAsu. The vector was present in DH5-alpha T1 phage resistant E. coli, resistant to ampicillin. An overnight culture was grown in 5 μL lysogeny broth (LB) and 100 μg/mL ampicillin. After 16 hours of inoculation in 37°C at 180rpm, minipreparation (miniprep) was performed to isolate the DNA, using the Promega Pureyield™ miniprep kit. During miniprep, 900 μL of pure ethanol (EtOH) was added, before the supernatant (containing DNA) was transferred to the microcolumn. DNA concentrations were measured using Biodrop UV/VIS spectrometry. [27]

0,5 μL of isolated DNA, expressing the EGFR gene, was transformed in 50 μL of competent cells. For competent cells, NEB competent (turbo) k12 strain E. coli were used, which were previously conserved in 50% glycerol stock at -80°C. For 30 minutes, the DNA/cell mixture was inoculated on ice. A heat-shock of 30 seconds 43.7°C was applied, after which the tubes containing the mixture were directly placed on ice for two minutes. After adding 450 μL of LB without ampicillin to the mixture, the tubes were placed in a shaking incubator at 37°C for 45 minutes. 15μL of 100ug/μL ampicillin was added to 10 cm LB agar plates. The culture and a 1/10 dilution were grown overnight at 37°C.

Of each plate, two colonies were picked and transferred to a tube containing 5 mL of LB stock and 100 μg/mL ampicillin. The tubes were grown overnight in 37°C at 180 rpm. Of each overnight culture, 500 μL was transferred to glycerol stock and stored in 80°C for later use. The remaining volume of the overnight culture was used for miniprep. The volumes of both the undiluted and the 1:10 diluted overnight culture were merged. Miniprep was completed using the Promega Pureyield™ miniprep kit, with the addition of 900 μL of ethanol before transferring the DNA to the microcolumn.

Polymerase chain reaction (PCR) was completed, using DNA with a concentration of 754.5 μg/μL, obtained from the Miniprep. The DNA was diluted to a concentration of 232.8 μg/μL, primers were diluted to a concentration of 10pmol/μL. The PCR reaction mixtures contained 9 μL diluted DNA, 0,5 μL reverse primer, 0,5 μL forward primer and 10 μL PCR master mix with Phusion polymerase. The following oligonucleotides were used as forward primer: 5'-GCG C CCA TGGG CTG CTG CAG GAG AGG G-3' with NcoI site (underlined). As a reverse primer 5'-GCG C AAG CTT ATG CTC CAA TAA ATT CAC TGC TTT GTG-3' with HindIII site (underlined).

The following thermocycling conditions were used for 30 cycles three-step PCR: Initial denaturation for 2 minutes at 98 °C, annealing at gradient temperatures for 30 seconds, extension at 72 °C for 30 seconds, followed by final extension at 72 °C for 5 minutes and hold at 4 °C. The following seven annealing temperatures were used in the first PCR reaction: 72 °C, 71.5 °C, 70.5 °C, 68.7 °C, 66.6 °C, 64.8 °C and 63.6 °C. The PCR product DNA was analyzed by gel electrophoresis in 1% agarose gel (Bio-Rad). A second polymerase chain reaction was completed, using a DNA concentration of 99.4 ng/μL. Instead of the previously described 7 annealing

temperatures, the amount has been reduced to the following five: 69°C, 68.6°C, 67.7°C, 66.1°C, 64.2°C.

For the third PCR the DH5-alpha T1 phage resistant E. coli, purchased from DNAsu with pJP1520 vector expressing both an ampicillin resistance and EGFR gene, was directly inoculated. Two cultures were grown overnight in 5 μL lysogeny broth (LB) and 100 μg/mL ampicillin. The DNA was isolated by miniprep. During miniprep, 900 μL EtOH was added to one of the two cultures before the supernatant was transferred to the microcolumn. Instead of transforming the DNA into competent cells, the isolated DNA was used directly in PCR. For PCR, the DNA concentrations were diluted to 48.62 ng/μL and 45.6 ng/μL for the samples treated with and without EtOH respectively. Three-step PCR was performed for 30 cycles, in equal thermocycling conditions as previously described. The annealing temperatures were set at 65,2°C and 64,2°C. Analysis was completed using gel electrophoresis in 1% agarose gel.

Before digestion, BL21 E. coli expressing pETM-30, previously stored in a 25% glycerol stock, were inoculated overnight. Miniprep was performed for isolation of pETM-30, with the addition of 100 μL EtOH during the process. Digestion mixtures of a total volume of 20 μL were produced for both the PCR product of the EGFR construct and pETM-30 DNA. Nco1 and HindIII were used as restriction enzymes. The digestion reaction mixtures were composed of 1μL of each restriction enzyme, 2μL of FastDigest Green Buffer. For pETM-30 3 μL 370.7 ng/μL DNA and 13μL autoclaved H_2O were used, whereas for the EGFR construct 5 μL 182.1 ng/μL DNA and 11 μL H_2O was added to the mixture. The mixtures were digested for three hours at 37°C, followed by ten minutes at 80°C.

The digested DNA was isolated by gel-purification after gel electrophoresis on 1% agarose gel. The Promega Pureyield™ kit was used for gel-purification of the DNA according to the protocol 'Wizard SV Gel and PCR Clean-Up System'. For ligation, a 7:1 mass ratio of vector DNA and insert DNA was used. 2 μL T4 DNA Ligase Buffer (10x), 1μL T4 DNA ligase, 11.6 μL of 3.526 ng/μL digested pETM30 and 5.3 μL 14.61 ng/μL digested EGFR kinase domain were incubated overnight. A negative control of plasmid self-ligation was included, which only contained vector DNA and not the insert DNA. The ligation process was terminated by heat inactivation at 65°C.

The ligation product was then transformed into 100 μL NEB competent (turbo) k12 strain E. coli following the previously described transformation protocol. 50 μg/mL Kanamycin was used as an antibiotic. Plates with only LB agar, LB agar and kanamycin and LB agar, kanamycin and the ligation product without insert DNA were included as negative controls and incubated overnight. Of the plate containing the ligation product of both the insert DNA and vector DNA, seven isolated colonies were picked. Colony PCR was completed for the different colonies. For PCR, red Taq polymerase and T7 forward and T7 reverse primers were used. The PCR reaction was completed in a three-step procedure: Initial denaturation for two minutes at 95 °C was followed by denaturation at 95 °C for 30 seconds Annealing took place at 65 °C for 40 seconds, followed by extension at 72 °C for 30 seconds. These steps were repeated for 30 cycles. The experiment was later repeated with 7 new colonies. During the second Colony PCR procedure, an additional

lysis step was included. Now, the isolated colonies were first denaturized for ten minutes in a water bath at 96°C.

Colony 1, 4 and 7 were grown overnight in 8 mL of LB media and 50 mg/mL kanamycin. Miniprep was performed according to the previously described protocol, with the addition of 900 μL EtOH. Of each colony, approximately 1000 ng of DNA was digested for two hours, following the previously described digestion protocol, using Nco1 and HindIII restriction enzymes. The digestion products were consecutively analyzed using gel electrophoresis in 1% agarose gel. In addition to digestion, PCR was completed with DNA of each overnight culture. Of each DNA sample, PCR was conducted with two different primers, using both T7 aspecific primers and previously described specific primers with Nco1 and HindIII cleavage sites. Taq polymerase was used for the extension. PCR was completed in a three-step procedure for 30 cycles: Initial denaturation at 95 °C for 5 minutes, denaturation at 95 °C for 30 seconds, followed by annealing at 65 °C for 40 seconds for the specific primers and 60°C for T7 forward and reverse primers and extension at 72 °C for 30 seconds. The analysis took place by gel electrophoresis in 1% agarose gel.

A DNA sample was sent to Eurofins for sequencing, containing 102.97 μg/mL DNA of colony seven and T7 reverse and T7 forward primers. The corresponding Seq. ID's are 20CF33 and 20CF34 for T7 reverse and T7 forward primers respectively.

3. Results

3.1 Amplification of EGFR construct

Miniprep was completed using an overnight culture of DH-5 alpha T1 phage resistant E. coli expressing EGFR in a pJP1520 vector. A DNA concentration of 3019 ng/μL was obtained. The vector was transformed into Turbo Competent E. coli. DNA concentrations of 268,4 ng/μL and 745,5ng/μL were isolated by miniprep for the undiluted and 1:10 diluted cultures respectively. Gradient PCR was completed using both the DNA of 745,5 ng/μL and 268,4 ng/μL. In figure 2, the electrophoresis gels of both PCR products are shown. Both gels show bands of approximately 3000 base pairs, at every annealing temperature. Also, thin bands around the 6000base pair (Bp) marker can be observed in gel 2B.

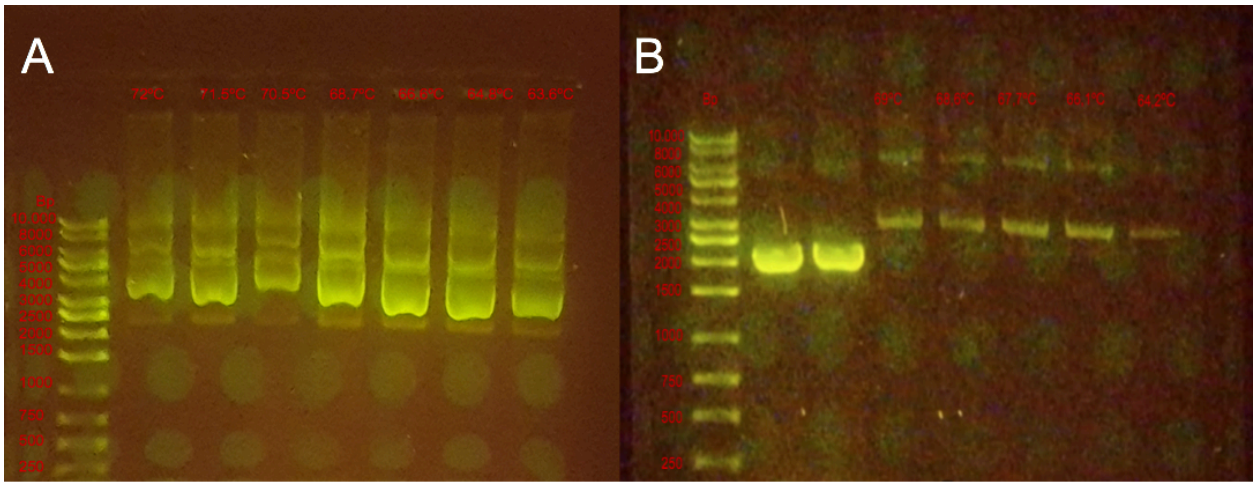

Figure 2: Gel electrophoresis after PCR of the transformed Turbo Competent E.coli. Lanes contain amplified DNA for different annealing temperatures in PCR. DNA of 745,5ng/μL and 268,4 ng/μL were used for PCR in gel A and B respectively. The bands in lane 2 and 3 in gel B, belong to a different experiment.

Miniprep was completed with overnight cultures of DH5-alpha T1 phage resistant E. coli purchased from DNAsu with pJP1520 vector expressing EGFR. DNA concentrations were obtained of 218.8 ng/μL and 205.4 ng/μL with and without additional ethanol in miniprep respectively. In figure 3, the results of an electrophoresis gel of both PCR products are shown. For all four samples loaded in the gel, clear bands can be observed around the 1500 base pair marker. No distinction can be made between samples treated with EtOH and those without EtOH addition during miniprep. Moreover, less bright bands are visible around 3000 Bp for all samples. The PCR products were merged, resulting in a DNA concentration of 182.1 ng/μL.

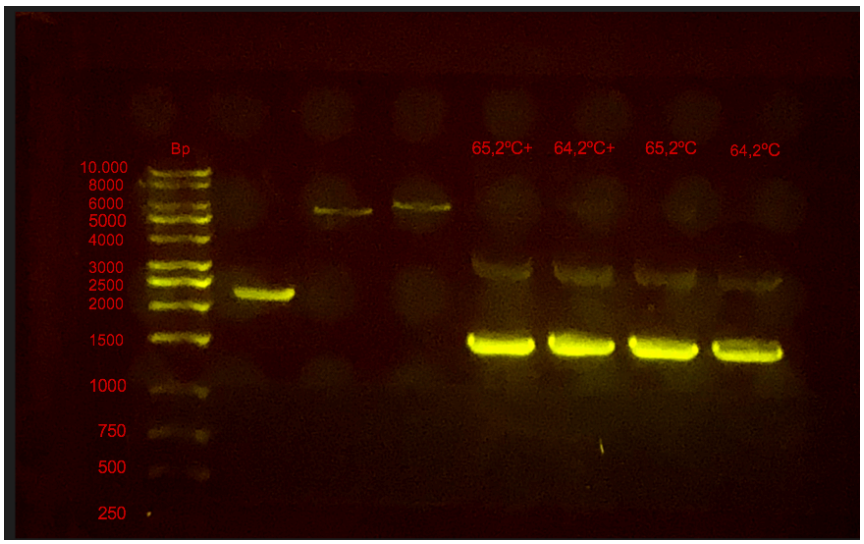

Figure 3: Gel electrophoresis after PCR with DNA isolated from DH5-Alpha T1 phage resistant E. coli purchased from DNAsu expressing EGFR. Samples treated with EtOH during miniprep are indicated by +. Concentrations of DNA used for PCR are 48.62 ng/μL and 45.6 ng/μL for the sample treated with and without EtOH respectively. Bands in lane 2 to 4 belong to a different experiment.

3.2 Construction of plasmid and transformation

Before digestion, a DNA concentration of 370,7 ng/μL was obtained from miniprep of an overnight culture of BL21 E. coli expressing pETM-30. Following, 1112.1 ng and 910 ng DNA of pETM-30 and the EGFR construct were digested respectively. In figure 4, the results of gel electrophoresis of the digestion products are shown. For the digestion product of EGFR (figure 4A), a clear band can be seen around 1500 Bp. Moreover, a less clear band can also be observed around 4000 Bp. For the digested pETM30 product, bands are shown around the 6000 Bp marker (figure 4B). Two samples of digested pETM-30 were obtained from gel purifications, resulting in concentrations of 2,389 ng/μL and 3,526 ng/μL. For EGFR, a concentration of 14.61 ng/μL was obtained from gel purification.

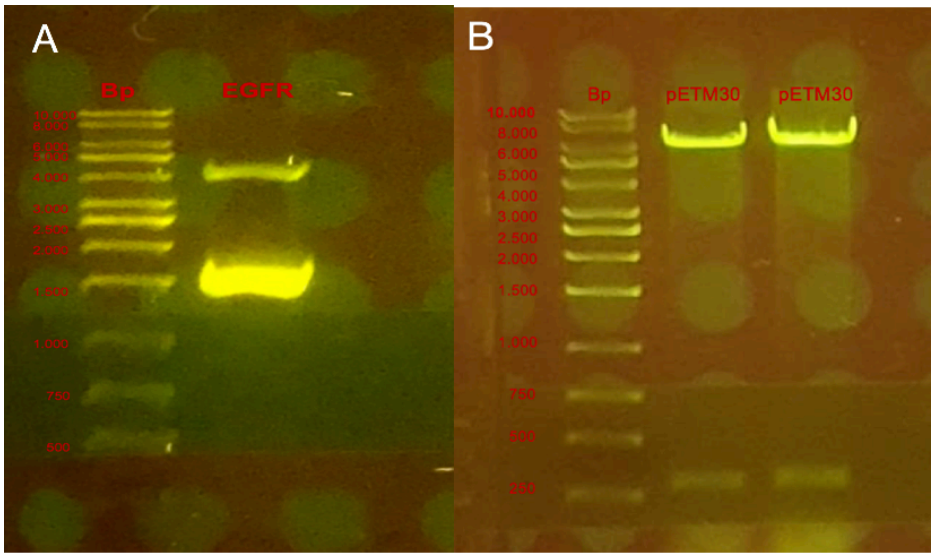

Figure 4: Gel electrophoresis of digestion products of EGFR (3a) and pETM30 (3b). Initially, 1112.1 and ng 910 ng DNA of pETM-30 and the EGFR construct were digested respectively.

The digestion products were purified and ligated using 'ready to go' T4 ligase. The ligation product was transformed in Turbo Competent E. coli. In figure 5, negative control plates are shown. In figure 5A bacterial growth can be observed after overnight incubation of both plates with LB agar and LB agar with kanamycin. On the negative control plate of transformation Turbo Competent E. coli with only vector DNA, without insert DNA, a single colony can be observed (figure 5B).

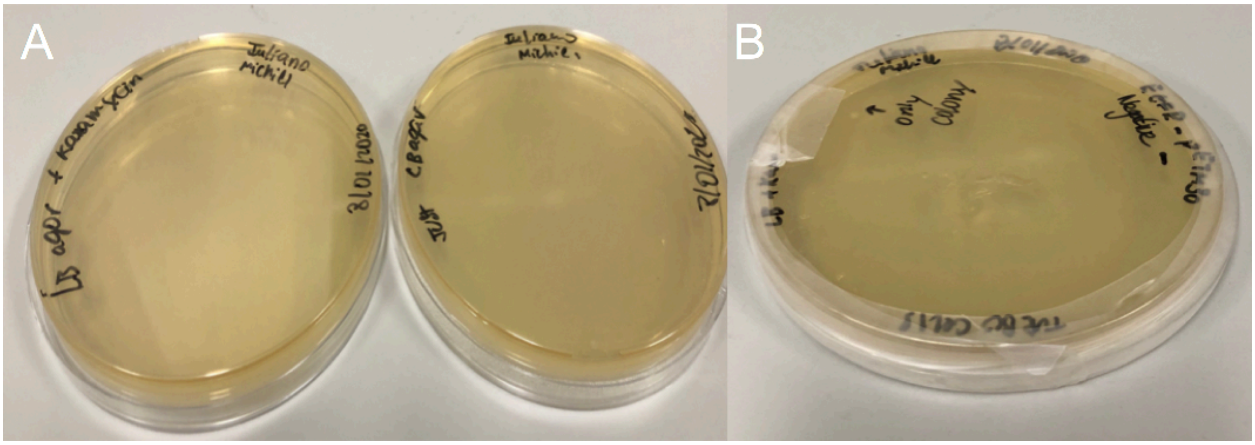

Figure 5: Negative control plates of transformation ligation product in Turbo Competent E. coli. Negative control plates in 3A of media quality, plates contained only LB agar media, left plate also contained 50 μg/mL kanamycin. Turbo Competent E. coli was transformed using only vector DNA without insert DNA, grown on LB agar in 50 μg/mL kanamycin in 3B.

The plate of the transformed ligation product in Turbo Competent E. coli can be seen in figure 6. Seven colonies (encircled in figure 6A) were used for Colony PCR. No PCR product could be observed by gel electrophoresis after the first Colony PCR. Seven new colonies were selected (figure 6N) and analyzed by gel electrophoresis after Colony PCR. Again, no bands were observable in the gel. Colonies 1,4,7 (figure 6N) were inoculated overnight.

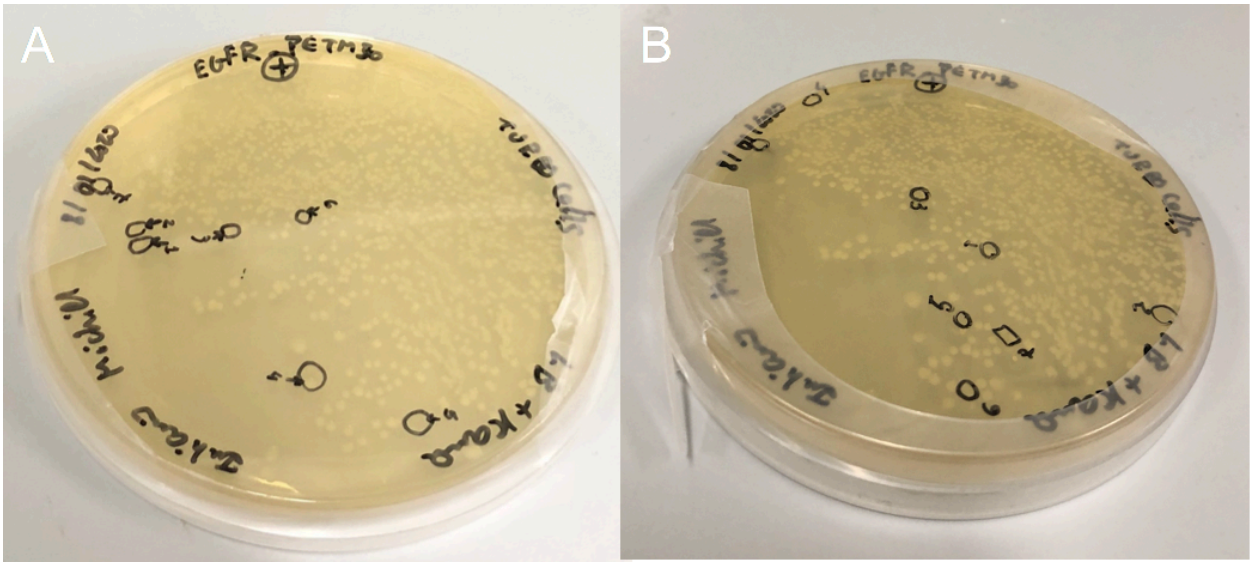

Figure 6: Plate of transformed ligation product in NEB competent (turbo) k12 strain E. coli grown on LB agar in 50 μg/mL kanamycin. Left plate encircled colonies used for the first Colony PCR, right plate encircled colonies used for second Colony PCR and overnight culture.

After overnight inoculation of colonies 1,4,7 (figure 6b), DNA was obtained after miniprep in concentrations of 633.9 ng/µL, 927.7 ng/µL and 1029.7 ng/µL for colonies 1, 4 and 7 respectively. Two-hour digestion was completed using Nco1 and HindIII as restriction enzymes. The analysis was completed by gel electrophoresis. In figure 7, bands can be observed at both 1500 bp and 6000 bp, for the digestion products of DNA from colonies 1, 4 and 7.

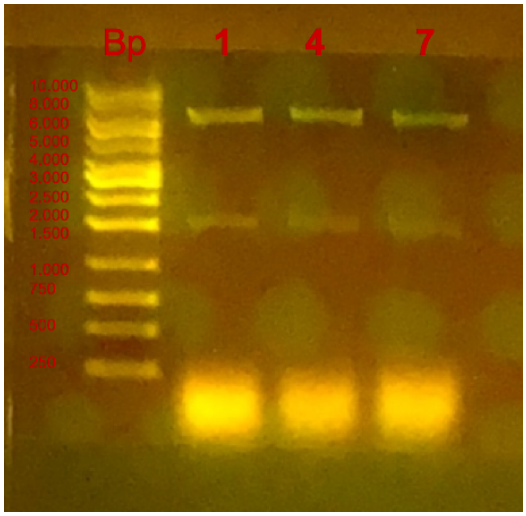

Figure 7: Gel electrophoresis after two hours of digestion of approximately 1000 ng DNA obtained from miniprep of inoculated colonies 1, 4 and 7. DNA concentrations of 633.9 ng/µL, 927.7 ng/µL and 1029.7 ng/µL were used for digestion for colonies 1, 4 and 7 respectively.

In figure 8, the results of an electrophoresis gel of PCR products of the Turbo Competent E. coli transformed with the ligation product are shown. For the PCR products using the specific primers, bands can be seen around 1500 bp for every colony. Also, thin bands can be observed around 3000 bp. For the PCR products of T7 primers, clear bands can be observed between 2000 and 2500 bp.

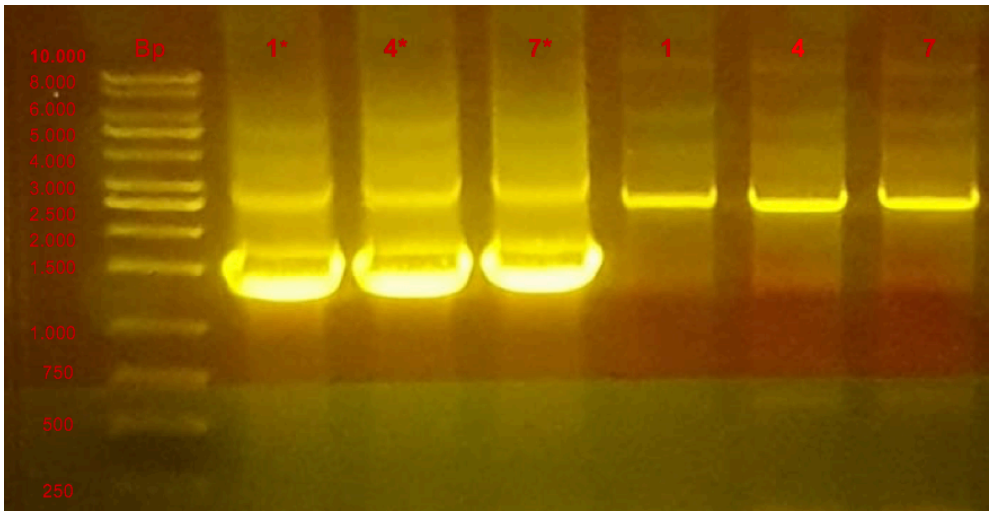

Figure 8: Gel electrophoresis of PCR-product after transformation of ligation product in Turbo Competent E. coli. During PCR for. 1, 4* and 7*, specific primers were used. For 1, 4 and 7, T7 primers were used in PCR. Taq polymerase was used for the extending of DNA, annealing temperatures were set at 65 ℃.*

4. Discussion

4.1 Design of EGFR construct

EGFR consists of 1211 amino acids. The extracellular domain is composed of residue 25 to 645, the transmembrane region ranges from residue 646 to 668 and the cytoplasmic domain comprises residue 669 to 1211, of which amino acids 683-958 host the catalytic domain of the kinase. [26] A gene encoding for residue 681 to 1211 was constructed, resulting in a 1590 Bp gene. The following primers were designed: 5'-GCG C CCA TGGG CTG CTG CAG GAG AGG G-3' as forward primers with NcoI site (underlined) and 5'-GCG C AAG CTT ATG CTC CAA TAA ATT CAC TGC TTT GTG-3' as reverse primers with HindIII site (underlined). The alignment of both primers is illustrated in figure 9. To both reverse and forward primers, non-complementary base-pair overhangs are added containing restriction sites resulting in sticky ends. The forward primer shows a melting temperature (T_m) of 64.46°C with a GC content of 70% and the reverse primer 65.28°C with a GC content of 37%.

Figure 9: Illustration of alignment of forward and reverse primer to EGFR gene in upper and lower figures respectively. The figure was created using snapgene.

4.2 Design of plasmid

A pETM30 plasmid was used for the cloning of the EGFR gene construct. The designed pETM30 plasmid with the EGFR gene construct is illustrated in figure 10. Undigested pETM-30 is composed of 6346 bp, whereas digestion with HindIII and NcoI leads to the removal of 347 bp. Ligation of the EGFR 1590 bp gene construct will result in a final plasmid size of 7589 bp. Various features of pETM-30 made it a suitable vector system for this study. However, other plasmids with similar properties, such as the pGEX-6P-1 used by Elloumi-Msedi *et. al. 2013*, could have also been used. [26] The pETM-30 plasmid and pETM expression vector series in general, have been widely used in this lab. Therefore, pETM-30 was the most suitable plasmid for this study, as it made for both the most readily available and economically viable option.

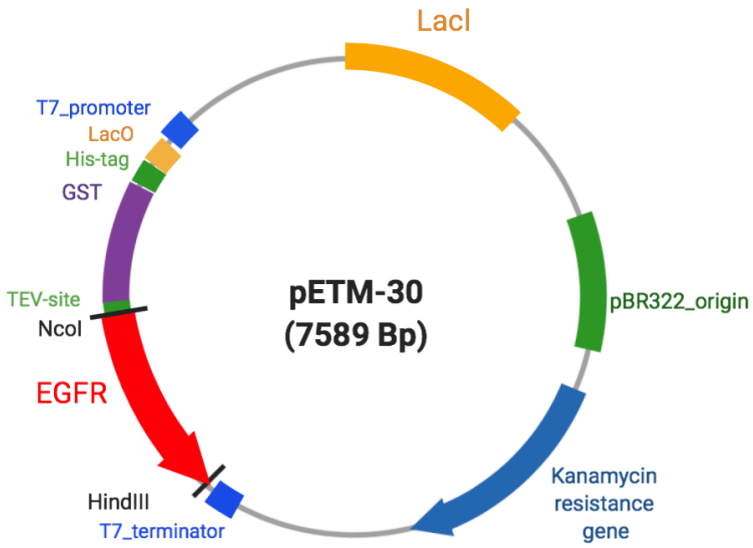

Figure 10: Overview of the designed pETM-30 Most relevant features of the plasmid are illustrated: Kanamycin resistance gene, LacI and lacO regulatory genes, EGFR gene construct, cloned in frame with the His-tag, GST and TEV-site, with NcoI and HindIII restriction sites and T7 promotor- and terminator-sites. The figure was created using Biorender.

The pETM expression vector series have been developed to increase the protein purification yield. Cloning of the EGFR kinase domain in pETM30 allows for the expression of a recombinant fusion protein. The expression is under the control of a T7 and lactose operon (lacO). In frame cloning of the EGFR kinase domain with the T7 promoter allows for selective induction of protein synthesis by the addition of either lactose or isopropyl β-d-1-thiogalactopyranoside (IPTG). Glutathione S-transferase (GST), Histidine-tag and a Tobacco Etch Virus nuclear-inclusion-a endopeptidase (TEV)-protease cleavage site are co-expressed and is directly fused to the EGFR protein construct. Furthermore, the selection of bacterial growth can be accomplished by the presence of a kanamycin resistance gene in the vector. [28,29]

The addition of a GST-tag to the recombinant protein greatly increases the solubility of the protein. Poor solubility of recombinant proteins is associated with a low purification yield. [30] GST-tags are able to dimerize on itself. Elloumi-Msedi *et. al. 2013* have previously shown that a dimerized GST fusion protein has allowed for EGFR kinase activity. Therefore, in this study it is particularly useful to co-express a GST-tag, for both the improvement of solubility and activity of the protein. [26]

The histidine-tag allows for a simplified purification process using immobilized metal affinity chromatography (IMAC). High affinity of histidine for metal ions, such as Cu^{2+}, Ni^{2+}, Co^{2+} and Zn^{2+}, leads to retention of the protein in the column, therefore separating it from other bacterial proteins present in the solution. [31] The elution of the protein can be accomplished by the addition of imidazole. The column has a greater affinity for imidazole. Therefore, the fusion protein is released from the column. Importantly, histidine tags are found to rarely alter protein characteristics and solubility. [30] The TEV-protease cleavage site can be cleaved by TEV-protease, releasing the EGFR kinase domain from the GST and His-tag fusion domains. However, in this study, the protein will not be cleaved, as a GST-induced dimerization is desired.

4.3 Amplification of EGFR construct

During miniprep, 900μL of 100% EtOH was added after the neutralization step to increase the yield, a protocol named Miraprep. In regular miniprep, DNA is isolated by spin column purification using commercial kits as Pureyield™. According to Pronobis et. al, the addition of ethanol after neutralization and before the washing step leads to precipitation of DNA on the silica column. Therefore, the yield does not only depend on the binding capacity of the column. This has shown to significantly improve DNA yield, without changing the quality of DNA. [27]

Phusion polymerase was used for plasmid cloning. Phusion polymerase is a high-fidelity polymerase and has improved accuracy compared to Taq polymerase. Previous studies have shown 50 times higher relative fidelity for Phusion polymerase compared to Taq polymerase. Taq polymerase, however, shows a superior speed and consequently reduces the extension time. Taq polymerase will be used during Colony PCR, in which the increased extension speed is particularly useful. In Colony PCR, only the size of the DNA construct is of importance. Therefore, mutations caused by the use of Taq polymerase are less relevant in this process and allows for improved speed. [32]

It was expected to see bands around the 1500 bp marker of the PCR product of transformed Competent Turbo E. coli with the JP1520 vector DNA, containing the EGFR gene. However, analysis by gel electrophoresis resulted in bright bands around the 3000 bp marker, as can be seen in figure 2. In figure 2A, it can also be observed that the gel was overloaded. Overloading of the gel leads to single bands being indistinguishable in the gel. Therefore, it was decided to repeat the PCR, using the same DNA isolated from the transformants. The concentration of PCR product DNA loaded in the gel was decreased from 232.8 ng/μL to 99.4 ng/μL. It can be seen that the bands were thinner and therefore more useful for analysis (figure 2B). However, bands were observed around the 3000 bp marker for the second PCR. The bands in the gel were

approximately double the size of the EGFR gene construct, which was expected to be present in the PCR product. The presence of a 3000 Bp construct, may be the result of self-annealing of primers, so-called self-dimerization. Here, primers are able to align and dimerize. Therefore, the dimer will be double the size of the initial construct. [33]

Premier Biosoft primer design guidelines suggest a ΔG of -9 kcal/mol or more positive as an acceptable value to prevent the formation of both the formation of self-dimers and cross dimers. A Cross-dimer ΔG of -9.89 kcal/mol was found by calculations under default reaction conditions, using NetPrimer, a free online tool offered by Premier Biosoft for primer design. Moreover, the ΔG self-dimer were -10.23 kcal/mol and -23.94 kcal/mol for the reverse and forward primer respectively. Especially the forward primer has a tendency of forming self-dimers. Therefore, it can be hypothesized that the bands observed at 3000 Bp are the result of the self-dimerization of the forward primer.

It has been previously described in literature that odds of primer-dimer formation increase with increasing concentrations of DNA. Das *et. al.* 1999 found that increasing the number of PCR cycles leads to an increased primer-dimer formation. In early PCR cycles, the amount of DNA doubles after each cycle. However, it was found that this exponential growth of PCR product eventually transitions into linear growth. At some point in the later cycles, the number of primers bound to the DNA template may exceed the amount of polymerase available for amplification. Therefore, primers bound to the template are free to react with other primer-template, especially when the designed primers have a high tendency of self-dimerization. The concentration of PCR products may start to decrease after this phase. [34]

Therefore, it was hypothesized that by decreasing the starting concentration of DNA, less primer-dimer formation would be observed. Accordingly, the starting concentration of DNA for PCR was, therefore, decreased from 99.4 ng/μL to 48.62 ng/μL and 45.6 ng/μL. As the quality of DNA may also influence the PCR product, it was decided to isolate JP1520 vector DNA with EGFR directly from the DH-5 alpha T1 phage resistant E. coli purchased from DNAsu. In addition, the effect of EtOH addition during miniprep on the quality of the DNA was examined. Consequently, two different DNA samples were produced, of both DNA isolated with and without the addition of EtOH during miniprep.

The gradient PCR has shown the clearest bands at an annealing temperature of 64.8°C (figure 2A). This matches the expectation, as the calculated T_m are 64.46°C and 65.28°C for forward and reverse primers respectively. New England Biolabs 'PCR protocol for Phusion High-Fidelity DNA Polymerase (M0530)' suggest using annealing temperatures equivalent to the T_m of the lower primer. [35] In PCR procedure following, the annealing temperatures were set at 65.2°C and 64.2°C to reach optimal DNA amplification.

In figure 3 it can be seen that bright bands can be observed around the 1500 bp marker for all samples. No distinction can be made between samples treated with and without EtOH during miniprep. Based on this gel, there is no reason to suspect alteration of the DNA quality by the use

of additional EtOH during miniprep. It is assumed, however, that the product does not contain pure monomeric product, as bands can also observable around 3000 bp with less clarity. Therefore, it is likely that the EGFR gene is both present as a monomer and dimer in the PCR product. In figure 3, it can be seen that there is a clear observable difference in the brightness of both bands at 1500 and 3000 Bp. The bands at 1500 Bp are far brighter than the bands at 3000 bp. Therefore, the concentration of the 1500 Bp DNA in the PCR product is likely to exceed the 3000 Bp DNA concentration. Before the EGFR gene is incorporated in the plasmid, restriction enzymes have to digest both ends of the gene. After digestion, only digestion products corresponding to 1500 Bp will be purified after gel electrophoresis. Therefore, it is hypothesized that impurities in the PCR product will not result in impurities of the final plasmid. Consequently, the PCR product containing both the monomeric and dimeric gene will be digested for the construction of the plasmid.

4.4 Construction of plasmid and transformation

In figure 4, the results of analysis by gel electrophoresis after digestion of both EGFR and pETM-30 can be observed. For the digestion product of EGFR, it was expected to see both bands at 1500 bp and 3000 bp. It was previously discussed that both monomeric and dimeric EGFR gene constructs were likely to be present in the PCR product. As the full PCR product was digested, it was likely to see both the monomer and dimer in the gel of the digestion product. In figure 4A it can be observed that a clear band is present at 1500 bp. Therefore, it can be assumed that the EGFR construct was successfully digested. Moreover, less bright bands can also be observed around the 4000 bp marker. This is larger than the projected impurity of 3000 bp corresponding to the dimerized gene construct. However, it can be also observed that the DNA ladder is suboptimal, as the 250 bp marker is faded. It can, therefore, be argued that the apparent 4000 bp band is, in reality, the dimer product. The nature impurity, in any respect, plays no role of significance in advances of this study. The suspected digested EGFR gene at 1500 bp will be cut and purified from the gel, therefore separating it from the impurity.

Digested pETM-30 is composed of 5999 bp. Consequently, it was expected for the digestion product of pETM-30 to show a band at 6000 bp in gel electrophoresis. In figure 4B gel electrophoresis indeed reveals a band at approximately 6000 bp. It was therefore assumed that both bands corresponded to digested pETM-30. Gel purification of pETM-30 resulted in a 5-fold lower concentration of DNA compared to gel purification of the suspected EGFR digestion product. For pETM-30, concentrations of 3.526 ng/μL and 2.389 ng/μL were obtained from gel purification, whereas for purification of the EGFR construct 14.61 ng/μL was obtained. A possible explanation underlying this variation in purification yield can be found in the use of additional EtOH in the purification of the EGFR construct and not for purification of pETM-30. As low concentrations were obtained from the purification of pETM-30, EtOH was added in a similar manner as the previously described Miraprep protocol, to improve the yield of the EGFR construct from purification.

After ligation by T4 DNA ligase, the ligation product was transformed in Turbo competent cells. In figure 5A, it can be observed that no bacterial growth was present for negative control plates

of media quality, containing both LB agar and LB agar with kanamycin. Therefore, it can be concluded that the quality of the media was sufficient. Furthermore, only a single colony could be observed after incubation of the negative control plate with only plasmid DNA transformed, without insert DNA (figure 5B). For the overnight culture containing transformants with both the plasmid DNA and insert DNA, a colony dense plate was obtained (figure 6). In case of a high quantity of self-ligation of the plasmid, meaning the plasmid closing on itself without incorporation of the EGFR gene, a similar amount of growth would be expected for both plates. The only plasmid feature necessary for growth is the kanamycin resistance gene, which is present in both plasmids with and without insert DNA. Consequently, it can be hypothesized that only a small fraction of plasmid DNA ligated on itself without the insert DNA.

To confirm successful transformation, Colony PCR was completed using seven isolated colonies (figure 6A). In Colony PCR, T7 reverse and forward primers were used, for amplification of the full recombinant EGFR construct, including the GST-tag, His-tag and TEV-site. The size of this recombinant gene construct is 2277 bp. Therefore, it was expected to see a band between 2000 and 2500 bp after analysis by gel electrophoresis. However, no observable bands were obtained after gel electrophoresis with the PCR colony product. During Colony PCR, thermal lysis was performed by heating the cultures to 96°C for 2 minutes. Therefore, it was hypothesized that lysis of the bacterial cells was incomplete, resulting in low starting concentrations of DNA template for PCR. Colony PCR was repeated, using seven new colonies (figure 6B). An additional thermal lysis step was included in the colony PCR protocol, in which PCR mixtures were placed in a 96°C water bath for ten minutes. However, despite the efforts of improving bacterial lysis, no result was obtained from the gel electrophoresis following Colony PCR.

As both Colony PCR reactions did not lead to observable results from gel electrophoresis, it was hypothesized that the amount of transformant DNA isolated was insufficient. Moreover, the number of cells in the PCR reaction was limited, as the culture was transferred directly from the plate into the PCR reaction mixture. In contrast, previous PCR reactions have been conducted using DNA isolated from overnight cultures, of which concentrations are significantly higher due to the replication of the bacteria. Consequently, it was decided to pick three random colonies to be inoculated overnight. The overnight cultures could then be used for isolation of DNA by miniprep. For all three cultures, the concentrations of isolated DNA were sufficient to reach approximately 1000 ng DNA in each PCR reaction mixture.

In figure 7, the results of the 2-hour digestion of the isolated DNA are shown. For digestion, HindIII and NcoI were used as restriction enzymes. Therefore, it was expected to see bands both around 1500 bp for the EGFR gene construct and around 6000 bp for the digested pETM-30 plasmid. In the case of self-ligation, it would be expected to only observe bands of 6000 bp, as the EGFR gene construct is not present. It can be seen that for all three colonies, both bands at 1500 bp and 6000 bp are present in the gel. Therefore, the digestion suggests the successful transformation of the pETM-30 vector with the ligated EGFR gene construct.

PCR was conducted for all three colonies, to obtain further confirmation of the successful transformation of the plasmid construct. For each colony, two different PCR reactions were completed. For the first PCR, the specific primers as designed in paragraph 4.1 were used, leading to an expected PCR product of 1590 bp. The second PCR was conducted using T7 forward and reverse primers, leading to the amplification of the full recombinant gene construct, including the GST-tag, His-tag and TEV-site of 2277 bp in total. In figure 8, gel electrophoresis shows bands around 1500 bp for the specific primers. Moreover, bands between 2000 and 2500 bp can be observed for the T7 primers. Therefore, it can be assumed with a high degree of certainty that the pETM-30 vector with the EGFR gene construct was successfully transformed in Turbo Competent cells.

4.5 Sequencing

A DNA sample was sent to Eurofins for sequencing, containing 102.97 µg/mL DNA of colony seven and T7 reverse and T7 forward primers. In figure 11, sequencing chromatograms of the first 50 base pairs using T7 reverse (11A) and T7 forward primers (11B) are shown. For the T7 reverse primer, a sequence of 802 bases was obtained, whereas the T7 forward primer resulted in a sequence of 1191 bases.

A clear chromatogram should contain well-formed peaks, in which single colors can be distinguished. Furthermore, the peaks should be evenly separated in the absence of background signals. The chromatograms in figure 11 do not apply to either of the previously stated requirements provided by Eurofins' guidelines. [36] Multiple sequence signals are present in both chromatograms, leading to overlapping peaks of variable heights. The sequencing, therefore, holds no value in the verification of the presence of the EGFR DNA construct in pETM-30.

Pre-diluted T7 primers were used in the samples prepared for sequencing. According to Eurofins' guidelines, multiple sequence signals can be the result of template and primer contamination, poor quality of template and primer or multiple priming of the DNA sequence. [36] Sequencing can, therefore, be improved by diluting fresh primers in nuclease-free water. Moreover, the preparation of three samples, including the DNA of colonies 1 and 4, will increase the odds of successful sequencing. Also, OD260/OD280 ratios can be determined, to ensure the use of sufficient DNA quality.

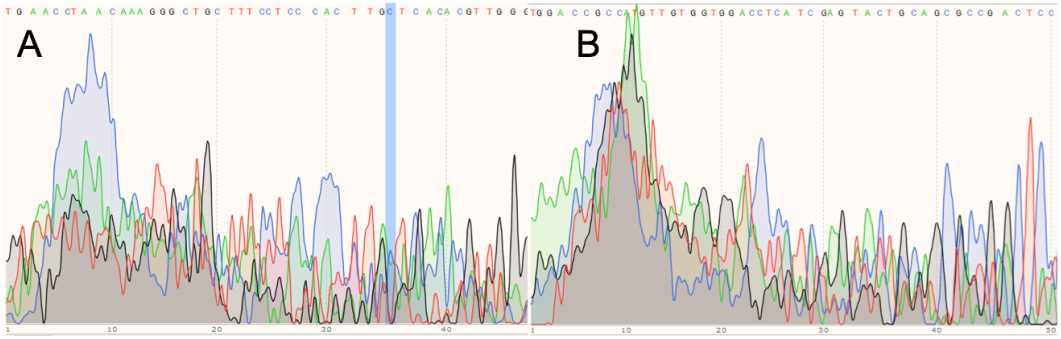

Figure 11: Sequencing chromatogram of the first 50 base pairs of samples containing 102.97 µg/mL DNA of colony seven using T7 reverse (9A) and T7 (9B) forward primers. The corresponding Seq. IDs are 20CF33 and 20CF34 for T7 reverse and T7 forward primers respectively.

5. Conclusion

During this study, primers were designed for the cloning of an EGFR gene construct, ranging from residue 681 to 1211. DNA of DH5-alpha T1 phage resistant E. coli expressing full human EGFR in a pJP1520 vector was isolated and used as template DNA for PCR. Analysis by gel electrophoresis indicated successful amplification of the EGFR construct by PCR, as bands were observed around the 1500 bp marker. The obtained DNA concentration of the PCR product was 182.1 ng/μL.

The PCR product and isolated pETM-30 DNA were digested by HindIII and NcoI. DNA concentrations of 14.61 ng/μL and 3.526 ng/μL were obtained from gel purification for the EGFR construct and pETM-30 respectively. Ligation was completed by ready-to-go T4 ligase in a 7:1 ratio for insert DNA and plasmid DNA respectively. The ligation product was transformed into Turbo Competent E. coli. The quality of growth media was assured by the absence of growth on the negative control plates. Moreover, negative control of transformation with the ligation product only digested plasmid DNA without insert DNA resulted in the growth of a single colony. Therefore, it was implied that self-ligation of the plasmid only occurred in moderate quantity, as for the full ligation product transformants a high colony density was obtained.

Initially, no results were obtained from Colony PCR, therefore three colonies were inoculated overnight and the was DNA isolated via miniprep. For all three colonies, digestion products showed bands at 1500 bp and 6000 bp after gel electrophoresis. Therefore, the digestion suggests the successful transformation of the pETM-30 vector with the ligated EGFR gene construct.

PCR was completed using both specific- and T7-primers. For all three colonies, gel electrophoresis bands of the PCR product using specific primers resulted in observable around the 1500 bp marker, matching the 1590 bp size of the EGFR gene construct. Furthermore, all PCR products of T7 primers were observed between 2000 and 2500 bp after gel electrophoresis. This has also matched the expectations, as the recombinant EGFR gene construct containing a GST-tag, His-tag and TEV-site contains 2277 bp in total. Based on PCR of transformed Turbo Competent E. coli, the EGFR gene construct has successfully been cloned in the pETM30 vector with a high degree of certainty. No useful results were obtained from sequencing. Sequencing, however, should be repeated in order to provide more evidence of successful plasmid cloning.

6. Future perspectives

During this research, the EGFR gene construct was successfully cloned in a pETM-30 vector, with a high degree of certainty. However, no satisfactory results were obtained from sequencing. Therefore, it would be logical to repeat sequencing in order to provide more evidence of successful cloning. New sequencing samples should be prepared with fresh primers and nuclease-free water. Furthermore, the DNA quality of all three transformants should be analyzed, for example by the calculation of OD260/OD280 ratios.

In case a successful ligation is confirmed by sequencing, the plasmids could then be transformed into bacteria strains specialized in protein synthesis. The BL21 E. coli strain has been well-established and is widely used for protein synthesis. Importantly, this strain has been genetically modified to optimize the production of proteins of recombinant nature. Deficiency in both OmpT protease and Lon protease increases the odds of both successful synthesis and purification. OmpT has a tendency of degrading the synthesized protein during purification, as it degrades extracellular proteins. Moreover, Lon protease specifically degrades foreign proteins such as the recombinant-EGFR protein. [28] Elloumi-Msedi *et. al.* have also used BL21 E. coli in their study for successful expression of both EGFR fused to GST as full human EGFR. [25,26] Therefore, the same strategy may be used going forward.

As the recombinant-EGFR construct contains a His-tag, the LOBSTR strain may also be an appropriate host for protein synthesis. LOBSTR cells are derived from BL21 and are also OmpT protease and Lon protease deficient. Furthermore, this strain has been genetically modified to improve His-tag affinity purification. This reduces the amount of contamination during the purification process. Contamination is likely to be caused by arnA and slyD proteins. Both these proteins contain Histidine-rich residues and may, therefore, be captured during immobilized metal affinity chromatography. However, LOBSTR cells express genetically modified arnA and slyD with reduced histidine residues and therefore lead to higher purity purifications after protein synthesis. [37]

After both successful synthesis and purification of recombinant-EGFR, the protein should be structurally characterized, for example by X-ray crystallography. Furthermore, the activity of the protein can be determined by protein incubation with [γ-^{32}P]ATP and successive Phosphorimaging, as previously shown by *Elloumi-Msedi et. al. (2013).*

Based on this research, no conclusion can be drawn regarding the validity and applicability of E. coli-based systems in the screening of inhibition by TKI's for various EGFR mutations. However, to the throughput of this method, various changes can be made to the cloning process. For instance, a new primer pair can be designed, the optimal primer-pair can then be used for all following expressions. This may result in faster cloning of EGFR, as changes of primer-dimer formation are reduced. Also continuously working with the same primer-pair, removes the need for gradient PCR, speeding up the process. Lastly, Colony PCR should be optimized as this saves time over an additional overnight culture and miniprep in the verification of successful cloning.

7. Literature

1. World Health Organization cancer. *12 september* https://www.who.int/news-room/fact-sheets/detail/cancer (2018).
2. Rotow, J. & Bivona, T. G. Understanding and targeting resistance mechanisms in NSCLC. *Nat. Rev. Cancer* **17**, 637–658 (2017).
3. Siegelin, M. D. & Borczuk, A. C. Epidermal growth factor receptor mutations in lung adenocarcinoma. *Lab. Investig.* **94**, 129–137 (2014).
4. Alberts, Bruce., Johnson, Alexander., Lewis, Julian., Morgan, David., Raff, Martin., Roberts, Keith., Walter, P. *Molecular biology of the cell*. (Garland Science, 2015).
5. Fabbro, D., Cowan-Jacob, S. W. & Moebitz, H. Ten things you should know about protein kinases: IUPHAR Review 14. *Br. J. Pharmacol.* **172**, 2675–2700 (2015).
6. Taylor, S. S. & Kornev, A. P. Protein kinases: Evolution of dynamic regulatory proteins. *Trends Biochem. Sci.* **36**, 65–77 (2011).
7. Modi, V. & Dunbrack, R. L. Defining a new nomenclature for the structures of active and inactive kinases. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 6818–6827 (2019).
8. Kobayashi, Y. & Mitsudomi, T. Not all epidermal growth factor receptor mutations in lung cancer are created equal: Perspectives for individualized treatment strategy. *Cancer Sci.* **107**, 1179–1186 (2016).
9. da Cunha Santos, G., Shepherd, F. A. & Tsao, M. S. EGFR mutations and lung cancer. *Annu. Rev. Pathol.* **6**, 49–69 (2011).
10. Jorge, S. E. D. C., Kobayashi, S. S. & Costa, D. B. Epidermal growth factor receptor (EGFR) mutations in lung cancer: Preclinical and clinical data. *Brazilian J. Med. Biol. Res.* **47**, 929–939 (2014).
11. Eck, M. J. & Yun, C.-H. Structural and mechanistic underpinnings of the differential drug sensitivity of EGFR mutations in non-small cell lung cancer. (2009) doi:10.1016/j.bbapap.2009.12.010.
12. Yang, Z. *et al.* Comparison of gefitinib, erlotinib and afatinib in non-small cell lung cancer: A meta-analysis. *Int. J. Cancer* **140**, 2805–2819 (2017).
13. Yun, C. H. *et al.* The T790M mutation in EGFR kinase causes drug resistance by increasing the affinity for ATP. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 2070–2075 (2008).
14. Russo, A. *et al.* Third generation EGFR TKIs in EGFR-mutated NSCLC: Where are we now and where are we going. *Crit. Rev. Oncol. Hematol.* **117**, 38–47 (2017).
15. Schil, P. E. Van, Hellmann, M. D., Peters, S. & Guidelines, E. Metastatic non-small cell lung cancer : ESMO Clinical Practice Guidelines for diagnosis , treatment Clinical Practice Guidelines. **29**, (2019).
16. Deng, X. & Nakamura, Y. Cancer Precision Medicine: From Cancer Screening to Drug Selection and Personalized Immunotherapy. *Trends Pharmacol. Sci.* **38**, 15–24 (2017).
17. Kobayashi, Y. *et al.* EGFR Exon 18 Mutations in Lung Cancer: Molecular Predictors of Augmented Sensitivity to Afatinib or Neratinib as Compared

with First- or Third-Generation TKIs. *Clin. Cancer Res.* **21**, 5305–5313 (2015).
18. Yasuda, H. *et al.* Structural, biochemical, and clinical characterization of Epidermal Growth Factor Receptor (EGFR) exon 20 insertion mutations in lung cancer (Science Translational Medicine (2014) 6, (225er1)). *Sci. Transl. Med.* **6**, (2014).
19. Russo, A. *et al.* Heterogeneous responses to epidermal growth factor receptor (EGFR) tyrosine kinase inhibitors (TKIs) in patients with uncommon EGFR mutations: New insights and future perspectives in this complex clinical scenario. *Int. J. Mol. Sci.* **20**, 1–20 (2019).
20. Hirano, T. *et al.* In vitro modeling to determine mutation specificity of EGFR tyrosine kinase inhibitors against clinically relevant EGFR mutants in non-small-cell lung cancer. *Oncotarget* **6**, 38789–38803 (2015).
21. Beau-Faller, M. *et al.* Rare EGFR exon 18 and exon 20 mutations in non-small-cell lung cancer on 10 117 patients: A multicentre observational study by the French ERMETIC-IFCT network. *Ann. Oncol.* **25**, 126–131 (2014).
22. Wu, S. G. & Shih, J. Y. Management of acquired resistance to EGFR TKI-targeted therapy in advanced non-small cell lung cancer. *Mol. Cancer* **17**, 1–14 (2018).
23. Chiba, M. *et al.* Efficacy of irreversible EGFR-TKIs for the uncommon secondary resistant EGFR mutations L747S, D761Y, and T854A. *BMC Cancer* **17**, 1–10 (2017).
24. Catalogue of Somatic Mutations in Cancer, version 90. https://cancer.sanger.ac.uk/cosmic/gene/analysis?ln=EGFR#variants.
25. Elloumi-Mseddi, J., Jellali, K., Villalobo, A. & Aifa, S. A premature termination of human epidermal growth factor receptor transcription in Escherichia coli. *Sci. World J.* **2014**, (2014).
26. Elloumi-Mseddi, J., Jellali, K. & Aifa, S. In vitro activation and inhibition of recombinant EGFR tyrosine kinase expressed in escherichia coli. *Sci. World J.* **2013**, (2013).
27. Pronobis, M. I., Deuitch, N. & Peifer, M. The Miraprep: A protocol that uses a Miniprep kit and provides Maxiprep yields. *PLoS One* **11**, 1–12 (2016).
28. Rosano, G. L. & Ceccarelli, E. A. Recombinant protein expression in Escherichia coli: Advances and challenges. *Front. Microbiol.* **5**, 1–17 (2014).
29. Dümmler, A., Lawrence, A. M. & de Marco, A. Simplified screening for the detection of soluble fusion constructs expressed in E. coli using a modular set of vectors. *Microb. Cell Fact.* **4**, 1–10 (2005).
30. Gräslund, S. *et al.* Protein production and purification. *Nat. Methods* **5**, 135–146 (2008).
31. Hengen, P. N. Purification of His-Tag fusion proteins from Escherichia coli. *Trends Biochem. Sci.* **20**, 285–286 (1995).
32. McInerney, P., Adams, P. & Hadi, M. Z. Error Rate Comparison during Polymerase Chain Reaction by DNA Polymerase. *Mol. Biol. Int.* **2014**, 1–8 (2014).
33. Chuang, L. Y., Cheng, Y. H. & Yang, C. H. Specific primer design for the

polymerase chain reaction. *Biotechnol. Lett.* **35**, 1541–1549 (2013).
34. Das, S., Mohapatra, S. C. & Hsu, J. T. Studies on primer-dimer formation in polymerase chain reaction (PCR). *Biotechnol. Tech.* **13**, 643–646 (1999).
35. BioLabs, N. E. No Title. *16 januari* (2020).
36. Results, E. Y. Dna sequencing results guide. *Eurofins Genomics* 1–14 (2014).
37. Andersen, K. R., Leksa, N. C. & Schwartz, T. U. Optimized E. coli expression strain LOBSTR eliminates common contaminants from His-tag purification. *Proteins: Structure, Function and Bioinformatics* vol. 81 1857–1861 (2013).