IN VITRO CONSTRUCTION OF EPIDERMAL GROWTH FACTOR KINASE DOMAIN

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

Abstract

Mutations of the kinase domain of epidermal growth factor receptor (EGFR) are a common cause of non-small cell lung cancer (NSCLC). Mutations lead to overactivity of EGFR resulting in increased proliferation, cell growth, cell survival and protein synthesis. Three generations of tyrosine kinase inhibitors (TKIs) have been developed, targeting various EGFR mutations. However, many mutational subtypes have been detected. TKI effectiveness is highly specific in regard to the mutation. Moreover, resistance to TKI is common, increasing the complexity of treatment selection. Optimal treatment for many rare mutations is yet to be determined. Therefore, optimal selection of TKIs could be achieved by the development of recombinant active EGFR tyrosine kinases in E. Coli. Binding of the various EMA approved TKIs to the mutated EGFR could be measured.

Supervisors:
Prof. Dr. Matthew Groves
Juliana F Vilachá

Michiel Zietse (3001830)
M.Zietse@student.rug.nl

29 november 2019

Table of Contents

1. Introduction

1.1 Cancer

In 2012, 1.59 million lung cancer-related deaths have been reported worldwide. [1] According to the World Health Organization, the annual death count has increased to 1.76 million in 2018, of which approximately 85% is related to tobacco use. [2] Besides smoking, other risk factors contribute to the development of lung cancer, including the exposure of the lungs to carcinogens such as asbestos, radon and arsenic. Moreover, air pollution, viral infections such as human papillomavirus (HPV), human immunodeficiency virus (HIV), Epstein-Barr virus, and diet seem to correlate with the development of lung cancer. [3]

Non-small-cell lung carcinoma (NSCLC) accounts for 85% of lung cancer cases. Small cell lung carcinoma (SCLC) accounts for the remaining 15%. Smoking is the leading cause of SCLC, which leads to the development of highly malignant tumors, with a worse prognosis compared to NSCLC. NSCLC can be classified according to originating cell type: adenocarcinoma (AC), squamous cell carcinoma (SCC) and large cell carcinoma. In NSCLC, the development of SCC is highly related to smoking. [3]

Besides the categorization of lung cancer based on originating cell type, a distinction can be made based on molecular subtype, as there are similarities in the disease manifestation within types with a similar genetic origin. Genetic changes may lead to the increased expression of oncogenes or decreased expression of tumor suppressor genes, both leading to increased cell division. Some of the major genes affected in NSCLC are Epidermal Growth Factor Receptor (EGFR), Kirsten rat sarcoma viral oncogene homolog (K-RAS), Echinoderm microtubule-associated protein-like 4 (EML4-ALK), Tyrosine-protein kinase MET (c-MET), Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha (PI3KCA), Vascular endothelial growth factor (VEGF) and B-cell lymphoma 2 (Bcl-2). [4]

The 5-year patient survival rate of stage IA NSCLC is 77-92% which decreases as stages progress. For patients with stage IIB, 5-year survival rate has dropped to 56%. In early-stage lung cancer, the treatment option of choice is surgery, with radiotherapy as a treatment option in patients in with a high perioperative risk profile. The effectiveness of a platinum-based chemotherapy in these stages of the disease remains undefined. [5]

In advanced-stage lung cancer, the untreated median survival is only 4 to 5 months. In this stage, the treatment of choice is directed at the molecular abnormalities that underly the disease. Various immunotherapeutics have been developed based on these targets. As in NSCLC activating EGFR mutations are most abundant, Tyrosine Kinase Inhibitors (TKI), that target signaling pathways downstream from the EGFR have been developed. At present, five TKI have been approved by the EMA for the treatment of NSCLC in patients with EGFR-positive mutations. Although these drugs have been shown to have a high efficacy, development of resistance is a major limitation in TKI treatment. [6]

2 Receptor tyrosine kinases

2.1 Physiological properties of receptor tyrosine kinases

Over 60 human Receptor Tyrosine Kinases (RTKs) have been discovered, consisting of approximately 20 subfamilies based on structural properties. The ErbB-family consists of four receptors: EGFR (ErbB-1/HER1), ErbB-2 (HER2), ErbB-3 (HER3), and ErbB-4 (HER4). RTKs are composed of three protein domains, a cysteine-rich extracellular ligand-binding domain, a transmembrane α-helix and an intracellular kinase domain. In the absence of ligands, RTKs are primarily present in monomeric form. The extracellular ligand-binding domain is able to bind signaling molecules, such as EGF. Receptor binding of epidermal growth factor (EGF) or transforming growth factor-alpha (TGF-α) leads to dimerization EGFR.

Unlike most RTKs, EGFR kinase domains are not activated by transautophosphorylation. Instead, EGFR is activated by an asymmetric dimerization process. Here, the C-lobe of one kinase domain connects with the N-lobe of the other kinase. [7] Dimerization leads to the generation of a binding surface for intracellular proteins, as a result of conformational changes. Tyrosine residues in the C-terminal tails are then phosphorylated by the N-lobe connected kinase domain.[8]

Signaling cascades result from phosphorylation of intracellular signaling proteins binding to the binding surface. EGFR is able to recruit different signaling proteins, as seen in figure 1. Src-Homology domain 2 (SH2) of the growth factor receptor-bound protein 2 (GRB2) selectively bind phosphotyrosines. Upon phosphorylation, the RAS/RAF/MEK/MAPK pathway is induced (fig 1). RAS is activated by the conversion of GDP to GTP, leading to downstream signals of the kinase cascade. Eventually, transcription factors including c-fos, c-jun, c-myc, NF-$\kappa\beta$, signal transducer, activator of transcription (STAT), zinc finger transcription factor and Ets family members, are activated. This leads to the transcription of various proliferative proteins, stimulating cell growth and proliferation and prolonging cell survival. Besides the RAS/RAF/MEK/MAPK pathway, other signaling cascades such as the PI3K-AKT-mTOR pathway may also contribute to proliferative gene transcription. [8,9]

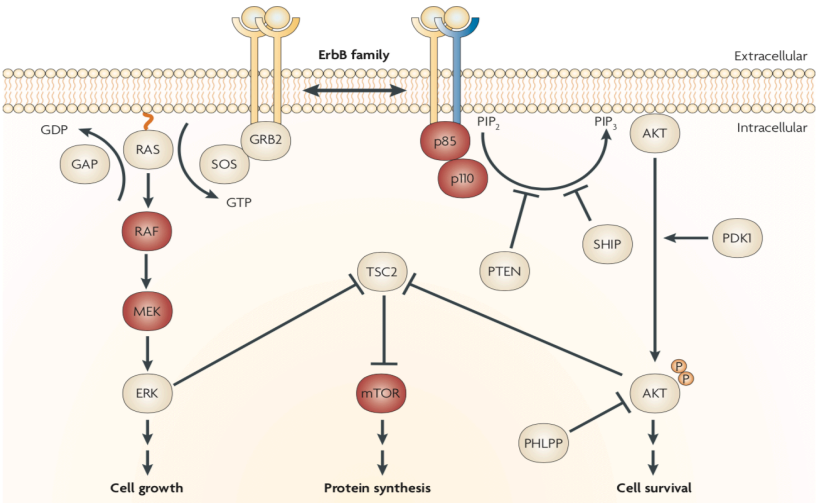

Figure 1: Schematic overview of downstream signals, by activation of epidermal growth factor receptor (part of the ErbB family). On the left, activation of the RAS/RAF/MEK/MAPk pathway, resulting in cell growth. On the right, activation of the PI3k-ATK-mTOR pathway. [6]

2.2 Structural properties of receptor tyrosine kinases

The kinase domain of RTKs is composed of a bilobal structure. A structural overview of the tyrosine kinase domain can be seen in figure 2. The N-lobe and the larger C-lobe are connected by a hinge region. In the cleft between the lobes, hydrophobic residues of the hinge region allow for binding of ATP by the formation of hydrogen bonds of adenine with the protein backbone. [10,11] The N-lobe is formed by 5-stranded β-sheets (β1-5) and an alpha helix (αC). One of the functional structures composed in the N-lobe is the P-loop. This is a well-conserved, flexible glycine-rich domain. An important role for the P-loop is the binding of ATP. The hydrophobic nature of the P-loop allows for positioning of the binding surface of the nucleotide, enabling the γ-phosphate of ATP to be used for catalysis. [12]

The αC-helix, belonging to the N-lobe, is also attached to the C-lobe. The αC-helix must be bound to the β-sheet core for the kinase to be able to efficiently perform catalysis. [12] In the activated state, the αC-helix moves towards the ATP-binding pocket (αChelix-in). A salt-bridge is formed between Lys^{745} in the β3 strand (AXK-motif) and Glu^{762} of the αC-Helix. This leads to the coupling of ATP's α- and β-phosphates to the αC-helix. Here, hydrogen bonds are formed between lysine side chains and the oxygen atoms of ATP. When inactivated, rotation of the αC-Helix (αChelix-out) leads to the conservation of the Glu residue. [12,13]

The N-terminal of the C-helix, located at the hinge, contains a gatekeeper residue. The ATP-binding site, back-pocket, is guarded by this residue. Access to the back-pocket depends on the bulkiness of the gatekeeper amino acid side chain. Gatekeeper mutations are common and have shown to play an important role in the development of TKI-resistance. [10]

The C-lobe is composed of mostly α-helices and also contains four β-strands (β6-9). Protein substrates are bound in the helices of the C-lobe. Therefore, the C-lobe plays an important role in the catalytic function of the kinase. The catalytic loop is enclosed in β6 and β7, whereas the DFG-motif, as a part of the activation loop, is located between β8 and β9. ASP in the catalytic loop is a base acceptor, playing a role in the proton transfer. The activation loop (A-loop), has an important regulatory function. The A-loop is able to modulate kinase activity by blocking substrate from binding the C-lobe. [10] In addition, the inactive conformation of the A-loop prevents rotation of the αC-helix towards the catalytic domain. [14]

The DFG-motif is composed of aspartate, glycine and phenylalanine. In the inactive conformation, the Asp residue is orientated away from the active site of the kinases (DFG-out). Here, the ATP-binding site is occupied by Phe. [13] In active state (DFG-in), the positions of Asp and Phe swap, a process known as 'DFG-flip'. [15] This enables aspartate to bind Mg^{2+} in the catalytic domain, positioning the γ and β phosphates correctly, to be used for phosphate transfer. Hydrophobic interactions are formed between Phe of the DFG-motif and hydrophobic residues of both the αC-helix and Y/HRD-motif of the catalytic loop. [10] This leads to the formation of a hydrophobic pocket, increasing the stability of the DFG-in conformation. [15]

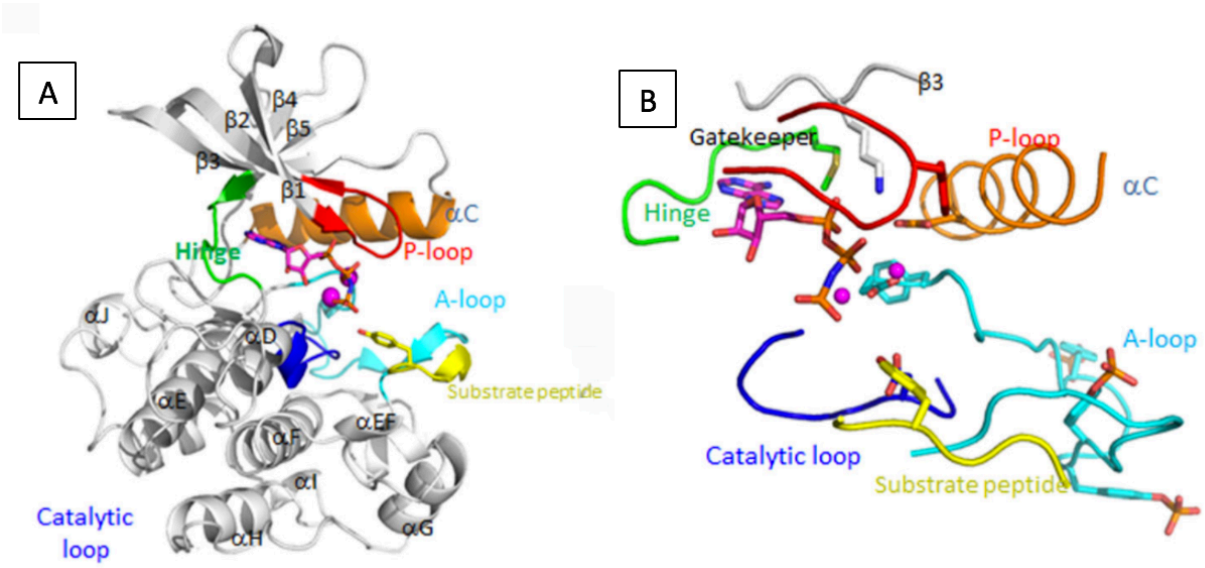

Figure 2: Structural overview of functional domains of protein kinases in active conformation. Front view of kinase domain (A) and a close-up of ATP active site (B). ATP molecule occupies ATP binding-pocket, connected to the hinge, Mg^{2+} ions displayed in magenta. Functional domains displayed in figure include the catalytic loop, A-loop, P-loop, αC-helix and the hinge. [10]

3. EGFR in non-small cell lung cancer

3.1 EGFR mutations

Activating mutations in the EGFR may result in the overexpression of proliferative factors, leading to unrestrained cell division and growth. [9] EGFR mutations are more commonly found in light- and never-smokers that develop NSCLC. For somatic EGFR mutations, the most frequent are in-frame deletions in exon 19 (Del19) (45%), exon 21 L858R mutation (40%) and in-frame insertions in exon 20 (Ins20) (5-10%).

L858R point mutations, the substitution of a single nucleotide in the DNA, lead to the transcription of arginine instead of leucine. The mutation occurs in exon 21 domain, encoding the A-loop. As previously described, the A-loop is an important regulator of kinase activity. In active conformation, the leucine is positioned on the proteins surface, whereas in inactive state the residue is packed in a hydrophobic pocket. Importantly, leucine is a hydrophobic amino acid, while arginine is a basic amino acid. Due to its polarity, the arginine side chain is unable to form hydrophobic interactions. Thus, the inactive conformation kinase domain is destabilized. The lowest energy state is reached when arginine is exposed on the protein's surface, where it is exposed to the solvent. This leads to the protein conformational equilibrium favoring activated state.[14] As a result, the receptor is now constitutive active, causing an increase of downstream signals. Eventually, the induction of multiple signaling kinases leads to the production of proliferative factors, resulting in cell growth and proliferation. [18]

Part of exon 19 encodes for the αC-helix. In case of exon 19 in-frame deletions, 3 to 8 residues are removed from the αC-helix. Therefore, the loop is shortened, which leads to the disfavor of the inactive helix-out conformation. The helix is, therefore 'pulled' into active conformation, leading to increased kinase activity. [19] Clinically over 30 subtypes of Del19 have been observed. [20] Besides these more commonly found somatic mutations, many other mutations have been described. [16] At present as much as 2794 unique mutations of the EGFR have been reported in the COSMIC database. [17]

In order to inhibit the cellular effects of increased EGFR signaling, three generations of Tyrosine Kinase Inhibitors (TKI's) have been developed. Inhibition of tyrosine kinase activity not only causes insensitivity to the effects of EGF and TGF-α, but also leads to the upregulation of proapoptotic proteins. This restores the balance of proapoptotic and antiapoptotic proteins, resulting in increased apoptosis of tumor cells. [16]

First-generation TKIs gefitinib and erlotinib are currently used as first-line treatment in advanced NSCLC in patients with an activating EGFR mutation. Both first-generation TKI's display a similar mode of action. Gefitinib and erlotinib show reversible binding of the RTKs in the ATP-pocket. [20] They are able to bind the kinase in either active or inactive conformation. Occupation of the ATP-pocket prevents the transfer of ATP phosphates and therefore results in inhibition of the receptor. [19]

Inhibition of EGFR by first-generation was proven to be a successful strategy in treatment of EGFR-mutated NSCLC. In fact, TKIs have resulted in superior progression-free survival (PFS) rates in NSCLC with EGFR mutations, compared to traditional chemotherapeutics. In the most common Del19 and L858R mutations, a PFS of 9-13 months has been demonstrated. [20] Both first-generation TKI present low binding affinities for the wild-type (WT) RTKs. For both common mutations, the binding of gefitinib and erlotinib results in a decreased affinity for ATP in the RTK. [16] Erlotinib has also shown efficacy in EGFR-wild-type NSCLC patients, whereas gefitinib is only effective in patients that harbor an activating EGFR mutation. In EGFR-mutant patients, no significant difference in the clinical efficacy of erlotinib and gefitinib has been observed. [21] However, in vitro studies suggest that the efficacy of erlotinib and gefitinib may differ between mutational subtypes (table 1). [20]

In 2013, afatinib has been approved by both FDA and EMA. This second-generation TKI binds irreversibly to Cys^{797} in the ATP-pocket of EGFR. The nucleophilic addition of sulfur allows for the formation of a disulfide bond, thus covalently binding cysteine. The covalently bound afatinib blocks the binding of ATP in the EGFR, preventing kinase activity. [19,22] In the LUX-lung 7 trial, improved PFS has been shown for afatinib compared to gefitinib in the first-line treatment of EGFR-mutant NSCLC. [17] However, afatinib only improved PFS in patients with Del19-, but not in L858R-tumors. [20] Also, afatinib showed a greater affinity for WT-EGFR, leading to a more severe toxicity profile. As a result, afatinib has not yet attained an undisputed position as the treatment of choice. [5]

In advanced EGFR-activating NSCLC, TKIs are considered the first-line treatment of choice. However, according to the January 2019 updated ESMO clinical practice guidelines, no consensus has been reached on which specific TKI is preferred drug. Either erlotinib, gefitinib and afatinib may be used in first-line treatment. [24] Nonetheless, in vitro studies have shown variable inhibitory responses to various EGFR-mutational subtypes (table 1).

Table 1 Sensitivity screening of different EGFR-mutations in mouse Ba/F3 cells for binding with various TKIs. IC50 values (nM) for TKI binding, <10, 10-99, 100-999 and >1000 in blue, light blue, yellow and red respectively. Results of In vitro studies summarized by Kobayashi et al. [20]

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	
EGFR wild type with interleukin-3			9350	>10 000	>100	>1000	>1000	3078	1549
Plasma drug concentration			(448–2717)	(2717–4040)	(69–130)	(166–238)	(N/A–132)	(400–600)	N/A–N/A

The concentration (nM) of added drug at which half of the maximal inhibition occurs (IC_{50}) has been determined for different TKIs for specific mutations (*table 1*). It can be observed that different subtypes of the Del19 mutation, respond differently to varying TKI treatments. For example, the IC_{50} value of gefitinib in Del19 subtype, delE746_S752insV, is 306 nM. In this mutation, a more favorable response is seen for erlotinib, with an IC_{50} of only 14nM. In contrast, in mutation-subtype delL747_P753insS a better response to gefitinib is observed. Importantly, even more Del19 subtypes have been discovered, some of which have yet to be screened for TKI binding. [20] In this in vitro study, afatinib has shown a lower IC_{50} value in every mutation compared to first-generation TKIs.

3.2 Resistance in TKI treatment

Inevitably, after an initial period of improved survival during treatment with either first or second-generation TKIs, the efficacy is diminished. Secondary mutations may lead to the development of resistance against first- and second-generation TKIs. A gatekeeper mutation T790M, in which a threonine is substituted for a bulkier methionine at position 790, leads to resistance in approximately 50% of the patients treated with afatinib, gefitinib or erlotinib. [19] Methionine causes steric hindrance, denying TKIs access to the ATP-binding site in the back pocket. However, an increasing affinity for ATP seems a more likely explanation of the occurrence of resistance in L858R primary mutations with an acquired T790M secondary mutation. Specifically, in L858R mutations, a T790M leads to increased ATP affinity (once more closely resembling that in WT-EGFR), whereas primary T790M mutations (without a previous L858R mutation) do not result in increased ATP affinity. [26]

Major resistance has been shown for the first-generation reversible TKIs due to increased ATP affinity, shifting the binding equilibrium in favor of ATP. Afatinib, being a covalent binder of the ATP-pocket, should be less affected by this shift in equilibrium. However, treatment with afatinib leads to epithelium-based toxicity before reaching substantial therapeutic effects in T790M, due to its potency against WT-EGFR. [27]

In vitro studies have shown that T790M mutations may both originate of cis- and trans-mutations. In case of trans mutations, heterogeneity of tumor cells exists. Here, multiple populations of tumor cells are present, each equipped with a different genotype. T790M mutations may be pre-existent in small populations. The populations with a germline T790M mutation have a survival advantage over the non-T790M mutated tumor cells during treatment with first- and second-generation TKs. Therefore, during treatment the cells may become the dominant population. This is clinically observed as TKI resistance via T790M. However, external stimuli and alteration of epigenetics may lead to de novo mutations. Here, previously T790M negative tumor cells acquire the T790M mutation. In these cis mutations, multiple mutations exist in the same cell population. [28–30]

Third-generation TKIs have been developed to specifically target T790M EGFR mutations, with osimertinib currently being the only EMA-approved TKI of this generation. Osimertinib irreversibly binds to Cys^{797}, inhibiting the kinase activity by the occupation of the ATP binding site. It also has proven efficacy in first-line treatment of L858R and Del19 mutations. [31]

The replacement of cysteine with serine (C797S) has been observed as a resistance mechanism against osimertinib. This replacement leads to the removal of the covalent binding site of osimertinib therefore causing resistance. However, C797S mutations have been shown to be sensitive to inhibition by first-generation TKIs. In addition, even C797S/T790M trans-mutations were found to be responsive to combined treatment with first- and third-generation TKIs. [20] Here the C797S population is responsive to first-generation TKIs and the T790M population can be treated with osimertinib. Previous studies have shown that in some cases, acquiring C797S paired with the conversion of $T790M^+$ to $T790M^-$ during treatment results in trans-mutation cell lines.

Therefore, combination therapy with different generation TKIs may be suitable in some cases of acquired resistance. [32]

Besides T790M, other secondary mutations have been observed that cause EGFR TKI resistance. Less common point-mutations, D761Y, T854A and L747S have all been shown to cause resistance to first-line TKIs. [33] D761Y has been observed after gefitinib treatment. Here, aspartate^{761} is substituted by tyrosine^{761}. This substitution occurs next to Glu^{762} of the αC-helix, which is able to form a salt-bridge with Lys^{745}, as previously described. How this mutation results in decreased sensitivity to EGFR TKIs, is yet to be determined. [34] In vitro sensitivity to third-generation of these L858R positive Ba/F3 cells with secondary D761Y, T845A and L747S mutations TKIs has been observed. Therefore, patients expressing these rare secondary mutations may benefit from treatment with osimertinib. [35]

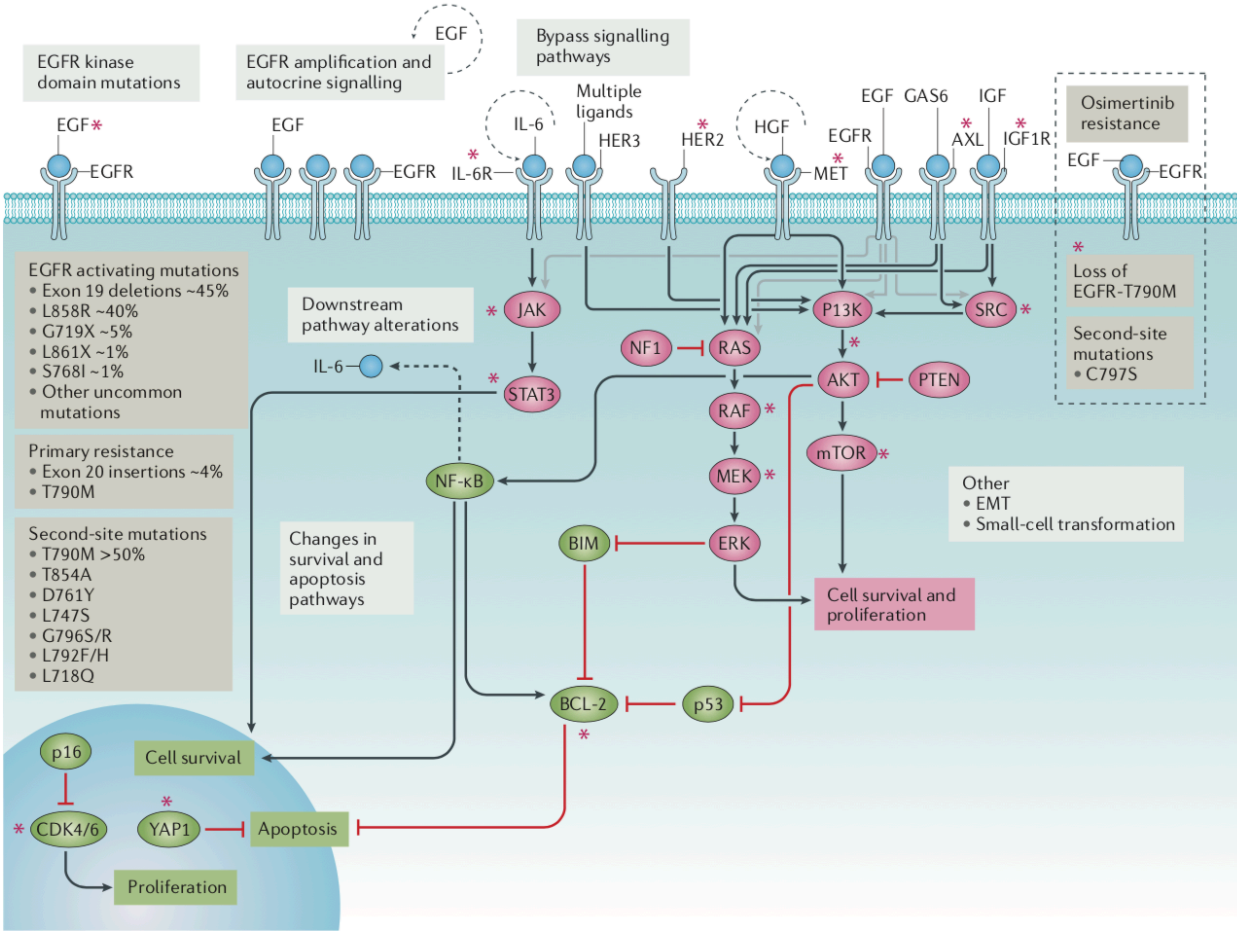

Figure 3: Overview of signaling pathways leading to resistance to TKI treatment in NSCLC with EGFR-mutations. TKI resistance can be the result of both primary and second-site EGFR mutations. EGFR amplification and autocrine signaling lead to the generation of more receptors and substrate. Bypassing of signaling pathways result in off-target resistance mechanism, avoiding the necessity of EGFR activation to reach increased cell survival and proliferation. Bypass signaling pathways illustrated in figure: MET amplification, HER2/3 amplification and activation of different receptors as insulin-like growth factor 1 receptor (IGF1R), AXL, Interleukin-6 receptor (IL6R).[30]

In addition to changes of the EGFR protein structure, other mechanisms of resistance have also been observed. An overview of mechanisms leading to TKI resistance can be seen in figure 3, the most frequent of which are MET amplification, HER2 amplification, BRAF mutations, NFκβ activation, transformation of NSCLC to SCLC and Epithelial to Mesenchymal transition (EMT). [31] Similar to EGFR, MET activation can lead to the induction of RAS/RAF/MEK/MAPK- and P13K-AKT-mTOR- pathways. Proliferation via both these pathways are also seen in resistance caused by increased signaling via IGF1R and AXL. Amplification of the receptor leads to an increased sensitivity to receptor-ligand. Activation and amplification of HER2/3 receptors lead to increased signaling via P13k-AKT-mTOR. These so-called 'off-target' resistance mechanisms are non-EGFR related and need a different approach than targeting TKI treatment. [30]

4. The complex reality of TKI treatment

More common mutations as Del19, L858R, T790M, and C797S have been studied extensively in order to find the TKI most suitable for treatment in the affected patients. In EGFR-mutated NSCLC, TKIs have shown great results, significantly improving PFS, compared to radiotherapy and non-specific chemotherapeutics. However, in reality, EGFR mutations come in many variations, each with its unique characteristics. A personalized approach is needed, in order to reach optimal treatment of EGFR-NSCLC.

For the development of a personalized treatment, it is necessary to define the tumor genomics. Genome analysis of EGFR-positive tumors can be accomplished by means of next-generation DNA sequencing (NGS). NGS has become an important tool and is used in many laboratories. Sequencing with NGS leads to the detection of both known and unknown mutations with high sensitivity. In addition, using NGS, a distinction can be made between cis- and trans-mutations. This can be detected by defining whether a mutation occurs in one or both alleles. In tumors, DNA needed for NGS may be isolated from circulating tumor DNA (ctDNA). [24,36]

This strategy, often referred to as liquid biopsy, may offer an improvement in treatment monitoring compared to repeated tissue biopsies. In liquid biopsy, circulating tumor DNA (ctDNA) can be prospectively isolated and analyzed. This approach offers a real-time genomic overview of the current tumor state. In contrast, tissue biopsy is a more static approach, which may reflect the tumor genotype of months ago. New mutations are generated continuously, of which detection is of great importance to the treatment strategy. Another limitation of tissue biopsy is the acquirement of tumor DNA via single site sampling. As previously discussed, heterogeneity may exist in tumors. Small populations of for example T790M positive tumor cells may not be included in the sample. However, resistance by trans-mutations have demonstrated sensitivity to multi-drug treatment of multiple generations TKIs. Therefore, detection is important to guide treatment. [24,36]

Advances in technology have made it possible to characterize tumors with great precision. However, a better understanding of the responses of different EGFR-mutations to various TKIs is needed in order to optimize treatment. Response to TKIs is highly specific to mutational subtypes. Even the most common somatic mutations differ in their response to TKIs. Previous studies have shown that the outcome of treatment with EGFR TKIs not only depends on the existence of Del19 (of which over 30 variants have been discovered[20]), but also on the specific deletional location mutation. [37] Possibly, the differences between Del19 mutation subtypes result in each subtype requiring a different TKI, or even conventional platinum doublet chemotherapy, to reach optimal treatment. [38]

For more rare mutations as EGFR Ins20, which account for 5-10% of EGFR mutations, responses to TKI are even less precisely defined. According to Kobayashi et al. 44 types of exon 20 insertions have been identified. [20] Many of these mutations have shown unresponsiveness to TKI treatment. Therefore, exon 20 insertions have been considered TKI resistant tumors. However, in vitro studies found that 10 to 20% of Ins20 variants can be inhibited by first-generation TKIs. [39] For other rare mutations, such as exon 18 deletions and insertions, variable sensitivity to TKIs has

also been observed. The importance of mutation-specific TKI screening in rare EGFR-mutations has been shown in previous studies. [25,40]

In the context of resistance resulting from a large variety of secondary mutations, the selection of the optimal TKI becomes more complex. As previously described, resistance to second-line treatment can lead to renewed responsiveness of patients with EGFR mutations to first-line treatment[20], or even to the disappearance of a specific point-mutation in a trans-mutational tumor cell population.[32] In vitro screening of these complex mutations may be rewarding. Secondary mutations D761Y, T845A and L747S mutations have shown in vitro sensitivity to osimertinib. [35]

Improved classification of different EGFR-mutations in order to predict their response to TKI treatment is needed. This is especially so in more rare cases and complex mutations, in which sensitivity to TKI treatment is often not well defined. In vitro screening can be an important tool for selecting the most effective TKI in a personalized approach. [20,41] In vitro models have shown to closely resemblance clinical observations in TKI pharmacodynamics. Furthermore, in vitro models to determine the therapeutic window, the pharmacokinetics of EGFR-TKIs have been developed. In vitro results have shown to match data of clinical trials. [42]

Optimal selection of TKIs could be achieved by the development of recombinant active EGFR tyrosine kinases based on patient-specific mutations. Binding of the various EMA approved TKIs to the mutated EGFR could be measured. Possibly, E. coli could be used for the synthesis of the receptors. Previously, a baculovirus system has been used to produce recombinant active EGFR. However, the development of an E. coli based system would drastically decrease screening costs and therefore hold the potential to play a role in guiding TKI treatment. [43]

5. Research goal

During this research, WT EGFR-kinase domain will be synthesized in Escherichia coli (E. Coli). The recombinant kinase domain can later be used for screening of the binding of EMA approved TKIs. The use of E. Coli reduces the use of cell lines of animal origin. The goal is to develop a fast screening method, which could be used to guide selection of optimal TKIs in patient-specific mutations. In addition, E. Coli could be used for the screening of new TKIs. The development of such an in vitro system could close the gap between pre-clinical knowledge and clinical decision-making.

All effector functions of tyrosine kinases originate from the intracellular domain and TKIs target the ATP-pocket of the EGFR, which is also part of the intracellular domain structure. Therefore, in order to measure binding affinities, only the intracellular domain of EGFR needs to be synthesized. Moreover, fusion of the recombinant tyrosine kinase with glutathione S-transferase (GST) allows for activity screening. In presence of ATP, the recombinant protein is able to dimerize. Dimerization activates the tyrosine kinase domain, allowing for phosphorylation. [43]

Expression of only the intracellular domain increases the chances of successful expression in the E Coli system, which is drastically reduced for proteins weighing above 60 kDa. [44] Also, the expression of transmembrane proteins is generally low. At times, even cell toxicity occurs, thought to be caused by the presence of the transmembrane domain. [45]

E. Coli is a well-established host of protein synthesis. The use of E. Coli and other bacteria species for protein synthesis have greatly reduced the use of animal tissue, especially in biochemistry. Apart from improved of animal well-being, there are more advantages of E. Coli systems for protein synthesis. Growth kinetics of E. Coli are unmatched, plasmid transformation is fast and easy, high-density cultures are easy to produce and growth media is cheap and readily available. [46]

Previous studies have successfully synthesized a recombinant EGFR-kinase domain using the BL21 strain. [43,47] Importantly, this strain is deficient in Lon protease. Lon protease degrades foreign proteins, thus decreasing the likelihood of recombinant protein synthesis. Moreover, the OmpT gene has been removed from the genome in this strain as OmpT degrades extracellular proteins, potentially degrading the synthesized protein during purification. [46]

The cDNA of the EGFR-kinase domain is purchased from DNASU. The cDNA will be amplified using polymerase chain reaction (PCR). Induction of the EGFR-kinase domain gene in E. Coli will be conducted using the pET-series vector. An important feature of the pET-series vector is the presence of a T7 promoter sequence. The EGFR-kinase domain is cloned in frame with T7 promoter. Recognition of the T7 promoter by T7 RNA polymerase (T7 RNAP) leads to transcription of the gene. Bacterial synthesis of T7RNAP can be induced by gene transfer with a second vector. The T7RNAP gene should be under the control of another promoter sequence, for example, lacUV5. Lactose and isopropyl β-d-1-thiogalactopyranoside (IPTG) activate the lacUV5 promoter. Addition of either lactose or IPTG will therefore initiate T7RNAP synthesis and the

resulting synthesis of the recombinant EGFR-kinase. In this manner, the protein expression can be controlled. [46]

In order to prevent the growth of plasmid-free cells, the vector contains a selection marker. The pET-series vector contains the bla gene. Product of this gene is β-lactamase, an enzyme leading to resistance of β-lactam antibiotics by hydrolysis of the β-lactam ring. Growth media can be treated with ampicillin, a first-generation ß-lactam antibiotic, resulting in the growth of ampicillin-resistance species only. [46]

Table 2 provides a chronological overview of the research project. In week 1, primers will be designed for transcription of EGFR. E. Coli expressing both the EGFR and bla gene are inoculated overnight in media containing ampicillin. Addition of ampicillin to the medium prevents growth of other non-vector containing bacterial species. Furthermore, overnight inoculation provides a sufficient density of bacteria for isolation of plasmid DNA using minipreparation (miniprep). In week 2, competent E. Coli (TurboCells) will be used for efficient transformation of the isolated plasmid DNA containing the EGFR gene. Here, the plasmid is replicated, providing enough copies for isolation. Miniprep will be used for isolation of DNA from TurboCells. This is followed by amplification of the DNA using PCR, generating sufficient number of copies to be integrated in the pET-series vector. [46]

In week 3, both the pET-series vector purchased from DNASU and the amplified DNA will be digested using restriction enzymes. During digestion, compatible overhangs are generated for insertion of the EGFR-kinase domain gene into the plasmid by ligation. TurboCells will be used for the transformation of the constructed vector. A single colony is transferred onto selective media. After successful transformation, growth of multiple colonies can be observed as a result of plasmid transfer. Screening of the constructed plasmid will be conducted using colony PCR. This is a high-throughput method, without the need for plasmid purification. Here, a colony is directly added to the PCR mixture. The PCR mixture contains primers specific for vector DNA. The PCR product can be analyzed by gel electrophoresis. [48]

After successful identification of the vector using colony PCR, full sequencing of the plasmid DNA will be conducted in week 4. In case the vector is proven to be successfully inserted, the vector will be transformed into multiple E. Coli strains. After incubation, the separate strains will be screened for the expression of the EGFR-kinase domain. In week 5, colonies expressing the EGFR-kinase can be used for protein production, initiated by the addition of IPTG. The EGFR-kinase domains will be isolated. The protein expressed from this construct can be screened for binding to EMA approved TKIs using biophysical methods.

Table 2: Gantt-chart providing overview of predicted timeline of the research-project.

Column1	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6
Design primers						
Overnight culture of EGFR						
Miniprep of vector						
Transformation on TurboCells						
Miniprep						
PCR						
Digestion of EGFR						
Digestion of Vector						
Ligation						
Transformation on TurboCells						
PCR colonies						
Sequencing						
Transformation on multiple E. coli strains						
Screening of expression						
Protein production						
Protein purification						
Result analysis						

Bibliography

1. Mao, Y., Yang, D., He, J. & Krasna, M. J. Epidemiology of Lung Cancer. *Surg. Oncol. Clin. N. Am.* **25**, 439–445 (2016).
2. World Health Organization cancer. *12 september* https://www.who.int/news-room/fact-sheets/detail/cancer (2018).
3. Akhtar, N. & Bansal, J. G. Risk factors of Lung Cancer in nonsmoker. *Curr. Probl. Cancer* **41**, 328–339 (2017).
4. West, L. *et al.* A novel classification of lung cancer into molecular subtypes. *PLoS One* **7**, 1–11 (2012).
5. Hirsch, F. R. *et al.* Lung cancer: current therapies and new targeted treatments. *Lancet* **389**, 299–311 (2017).
6. Sharma S, Bell D, Settleman J, Haber D. Epidermal growth factor receptor mutations in lung cancer. Nature Reviews Cancer 169-181, (2007).
7. Zhang, X., Gureasko, J., Shen, K., Cole, P. A. & Kuriyan, J. An Allosteric Mechanism for Activation of the Kinase Domain of Epidermal Growth Factor Receptor. *Cell* **125**, 1137–1149 (2006).
8. Alberts, Bruce., Johnson, Alexander., Lewis, Julian., Morgan, David., Raff, Martin., Roberts, Keith., Walter, P. *Molecular biology of the cell*. (Garland Science, 2015).
9. Normanno, N. *et al.* Epidermal growth factor receptor (EGFR) signaling in cancer. *Gene* **366**, 2–16 (2006).
10. 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).
11. Cowan-Jacob, S. W. Structural biology of protein tyrosine kinases. *Cell. Mol. Life Sci.* **63**, 2608–2625 (2006).
12. Taylor, S. S. & Kornev, A. P. Protein kinases: Evolution of dynamic regulatory proteins. *Trends Biochem. Sci.* **36**, 65–77 (2011).
13. 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).
14. Kumar, A., Petri, E. T., Halmos, B. & Boggon, T. J. Structure and clinical relevance of the epidermal growth factor receptor in human cancer. *J. Clin. Oncol.* **26**, 1742–1751 (2008).
15. Palmieri, L. & Rastelli, G. Ac Helix Displacement As a General Approach for Allosteric Modulation of Protein Kinases. *Drug Discov. Today* **18**, 407–414 (2013).
16. 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).
17. Catalogue of Somatic Mutations in Cancer, version 90. https://cancer.sanger.ac.uk/cosmic/gene/analysis?ln=EGFR#variants.
18. da Cunha Santos, G., Shepherd, F. A. & Tsao, M. S. EGFR mutations and lung cancer. *Annu. Rev. Pathol.* **6**, 49–69 (2011).
19. 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.
20. 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).
21. 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).
22. Wecker, Helga,. Waller, C. F. *Small molecules in Oncology*. (Springer International Publishing, 2018).
23. Park, K. *et al.* Afatinib versus gefitinib as first-line treatment of patients with EGFR mutation-positive non-small-cell lung cancer (LUX-Lung 7): A phase 2B, open-label, randomised controlled trial. *Lancet Oncol.* **17**, 577–589 (2016).
24. 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).
25. 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).
26. 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).
27. Ko, B., Paucar, D. & Halmos, B. EGFR T790M: Revealing the secrets of a gatekeeper. *Lung Cancer Targets Ther.* **8**, 147–159 (2017).
28. Hata, A. N. *et al.* Tumor cells can follow distinct evolutionary paths to become resistant to epidermal growth factor receptor inhibition. *Nat. Med.* **22**, 262–269 (2016).
29. Signor, S. A. & Nuzhdin, S. V. The Evolution of Gene Expression in cis and trans. *Trends Genet.* **34**, 532–544 (2018).
30. Rotow, J. & Bivona, T. G. Understanding and targeting resistance mechanisms in NSCLC. *Nat. Rev. Cancer* **17**, 637–658 (2017).
31. 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).
32. Thress, K. S. *et al.* Acquired EGFR C797S mutation mediates resistance to AZD9291 in non-small cell lung cancer harboring EGFR T790M. *Nat. Med.* **21**, 560–562 (2015).
33. 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).
34. Balak, M. N. *et al.* Novel D761Y and common secondary T790M mutations in epidermal growth factor receptor-mutant lung adenocarcinomas with acquired resistance to kinase inhibitors. *Clin. Cancer Res.* **12**, 6494–6501 (2006).
35. 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).
36. Deng, X. & Nakamura, Y. Cancer Precision Medicine: From Cancer Screening to Drug Selection and

Personalized Immunotherapy. *Trends Pharmacol. Sci.* **38**, 15–24 (2017).
37. Chung, K. P. *et al.* Clinical outcomes in non-small cell lung cancers harboring different exon 19 deletions in EGFR. *Clin. Cancer Res.* **18**, 3470–3477 (2012).
38. Kaneda, T. *et al.* Possible differential EGFR-TKI efficacy among exon 19 deletional locations in EGFR-mutant non-small cell lung cancer. *Lung Cancer* **86**, 213–218 (2014).
39. 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).
40. 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).
41. 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).
42. 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).
43. 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).
44. Gräslund, S. *et al.* Protein production and purification. *Nat. Methods* **5**, 135–146 (2008).
45. Horn, M. A. & Walker, J. C. Chapter 37 Expression and Assay of Autophosphorylation of Recombinant Protein Kinases. *Methods Cell Biol.* **49**, 531–541 (1995).
46. Rosano, G. L. & Ceccarelli, E. A. Recombinant protein expression in Escherichia coli: Advances and challenges. *Front. Microbiol.* **5**, 1–17 (2014).
47. 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).
48. Bergkessel M, Guthrie C. Colony PCR. Methods in Enzymology. 299-309. (2013)