Keivn Doornhof – S2706288

A computational approach to predicting drug resistance in kinase dependent lung cancer

Abstract

The use of kinase inhibitors in cancer treatment has shown much potential in improving how common variants of lung cancer are treated (Gautschi et al., 2012; Han et al., 2005; Shaw et al., 2013). The major issue all current inhibitors face is that after a period of treatment, resistance against the drugs tends to develop. A common mechanism for drug resistance in kinase driven lung cancer is the emergence of secondary mutations (Barouch-Bentov & Sauer, 2011). In some cases, it is possible to switch to a different inhibitor after resistance develops (J. C.-H. Yang et al., 2017). While this is often known for the most common resistance mutations, in cases with novel mutations there is often little to no information available to make informed clinical decisions. Computational methods such as homology modeling and molecular docking could provide a solution to this issue, but large scale implementation of such methods is difficult as it currently requires a large time investment by knowledgeable experts to interpret the relevant data (van Kempen et al., 2018). To help resolve this issue, molecular docking filters were designed that can identify docking poses that provide indicators of drug sensitivity in an automated manner. Validating these filters using known resistance mutations showed that such a method has the potential to drastically reduce the workload required to draw conclusions from docking data, potentially allowing for broader application of computational methods in predicting drug efficacy and informing clinical decision-making.

Introduction

Non-small cell lung cancer (NSCLC) causes around 1.6 million deaths worldwide each year, making it one of the leading causes of death in the western world (Heron, 2018; Wasserman, 2015). Based on cell size, the disease can be categorized into small-cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC). Out of these two types, NSCLC is by far the most common, making up around 85% of all cases (Hammerschmidt & Wirtz, 2009).

Like all forms of cancer, NSCLC is the result of uncontrolled cell proliferation. A common mechanism by which this occurs in NSCLC as well as other cancer types, is deregulation of kinase signaling (Paul & Mukhopadhyay, 2004). Kinases are involved in the regulation of many cellular processes including cell-growth, differentiation and apoptosis (ZHANG & LIU, 2002). Knowledge about the expression of mutated kinases in tumors is of high importance, as it often warrants the use of specific drugs to better treat the patient (Singh, Singh, & Silakari, 2016; Thomas, Rajan, & Giaccone, 2012). Kinases that are most commonly mutated in NSCLC to form oncogenic drivers are the epidermal growth factor receptor (EGFR), v-Raf murine sarcoma viral oncogene homolog B (BRAF) and Anaplastic Lymphoma Kinase (ALK). Taken together, mutations in these kinases drive an estimated 19-30% of all NSCLC cases (Kris et al., 2011; Vijayalakshmi & Krishnamurthy, 2011).

EGFR

Out of the tree kinases mentioned previously, the epidermal growth factor receptor is the most common driving factor in the development of lung cancer (Kris et al., 2011; Vijayalakshmi & Krishnamurthy, 2011). EGFR is a member of the ErbB family of kinases and consists of an extracellular domain containing the ligand binding site, a membrane-spanning hydrophobic domain and a cytoplasmic domain containing the tyrosine kinase domain (Ferguson, 2008).

In wild-type EGFR, the binding of an agonist such as epidermal growth factor (EGF) to the extracellular domain activates EGFR by causing dimerization and a structural change in the intracellular tyrosine kinase domain (Figure 1) (Lemmon & Schlessinger, 2010; X. Yu, Sharma, Takahashi, Iwamoto, & Mekada, 2002). Once activated, specific tyrosine residues are phosphorylated which can recruit proteins with Src homology 2 (SH2) and phosphotyrosine binding (PTB) domains (Normanno et al., 2006). These adaptor proteins can activate several signaling pathways such as MAPK and AKT, which are involved in regulating cell growth, proliferation and cell survival (Figure 2). Because of this, improper activation of EGFR can be a driving factor in cancer (Normanno et al., 2006; Oda, Matsuoka, Funahashi, & Kitano, 2005).

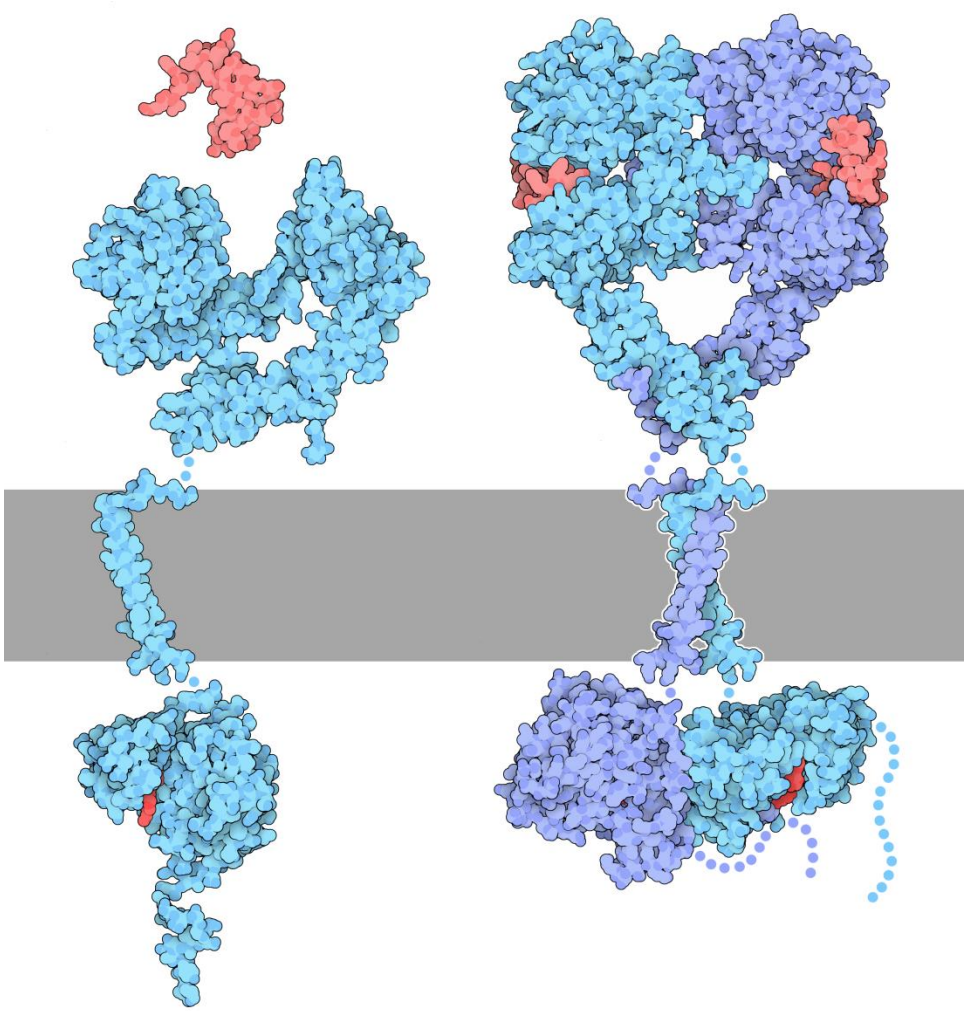


Figure 1 - Schematic view of EGFR (shown in blue and purple) dimerization after binding of Epidermal growth factor (ligands are displayed in red). The cell membrane is displayed as a gray bar (adapted from: Goodsell, 2010**).**

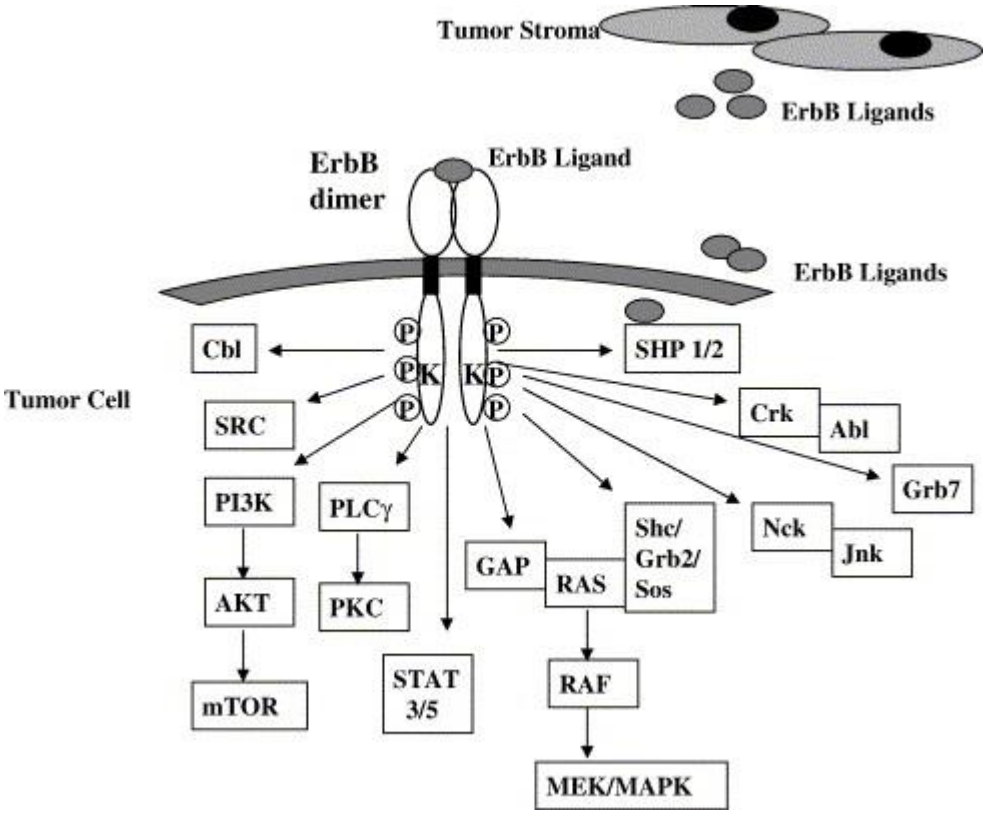


Figure 2 - ErbB activated pathways responsible for the oncogenic effects of uncontrolled EGFR activation (adapted from Normanno et al., 2006**)**

Deregulation of EGFR function can occur due to several reasons. These include increased EGFR expression due to gene amplification, mutations activating EGFR or a combination of the two. Activating mutation of EGFR consist primarily of exon 19 deletions and point mutations in exon 21 (Yoshida et al., 2007). Such mutations cause a change in the equilibrium between catalytically active and inactive EGFR, favoring bioactive conformations (A. Kumar, Petri, Halmos, & Boggon, 2008). In the past decade, much effort has been made to develop drugs that can prevent aberrant EGFR signaling. For this purpose, the EGFR tyrosine kinase inhibitors erlotinib, gefitinib, dacomitinib, afatinib and osimertinib were developed. These drugs target the ATP binding site on the tyrosine kinase domain, preventing ATP from binding and thus halting the phosphorylation of EGFR targets (Cakar & Göker, 2016). In clinical trials, such drugs were shown to provide a clear survival advantage over conventional chemotherapy when applied selectively to patients with EGFR driven NSCLC (Rosell et al., 2012; J. C.-H. Yang et al., 2017). In this patient group, these EGFR tyrosine kinase inhibitors (TKIs) are FDA approved and are currently used in clinical practice.

ALK

Another kinase that forms a common driving factor in NSCLC is Anaplastic Lymphoma kinase (ALK). ALK is believed to have a physiological function in the brain development during embryogenesis (Bayliss, Choi, Fennell, Fry, & Richards, 2016). Contrary to EGFR which is expressed in healthy tissue, ALK serves no physiological function in in adult human lung tissue and is normally not expressed here (Uhlen et al., 2015). Since ALK expression is limited in healthy tissue, ALK is a drug target that can be used to target cancer cells expressing it in a selective manner. In cancer, ALK expression is commonly the result of gene rearrangement leading to its expression as a fusion gene (Roskoski, 2013). ALK fusions were first identified in anaplastic large-cell lymphoma (ALCL) where it is mostly expressed as

a fusion with nucleophosmin (NPM). Currently, over 20 different ALK fusions have been reported in different types of cancer (Shaw & Engelman, 2013). In NSCLC, ALK is commonly expressed as a fusion with Echinoderm microtubule-associated protein-like 4 (EML4) (Zeng & Feldman, 2016).

ALK fusion proteins show constituent activity and are oncogenic due to their ability to activate signaling pathways involved in regulating cell survival and proliferation. Pathways influenced by ALK-fusion proteins can be seen in Figure 3 and include PI3K-AKT, JAK-STAT and MAPK (Roskoski, 2013). Studies have shown that the growth of tumors expressing ALK tends to be dependent on its continued expression (Gerber & Minna, 2010). To take advantage of this, ALK inhibitors can be used to treat tumors expressing ALK-fusion proteins. In the past decade, multiple such inhibitors have been developed and are now used in clinical practice. In clinical studies, ALK inhibitors were shown to greatly benefit patients with ALK-positive lung cancer compared to chemotherapy treatment (Roskoski, 2017; Rothenstein & Chooback, 2018; Shaw et al., 2013). For ALK inhibition, the inhibitors crizotinib, ceritinib, alectinib, brigatinib are currently FDA approved. Another inhibitor, entrectinib, is currently under FDA review and could be added to this list in the future (Drilon et al., 2017).

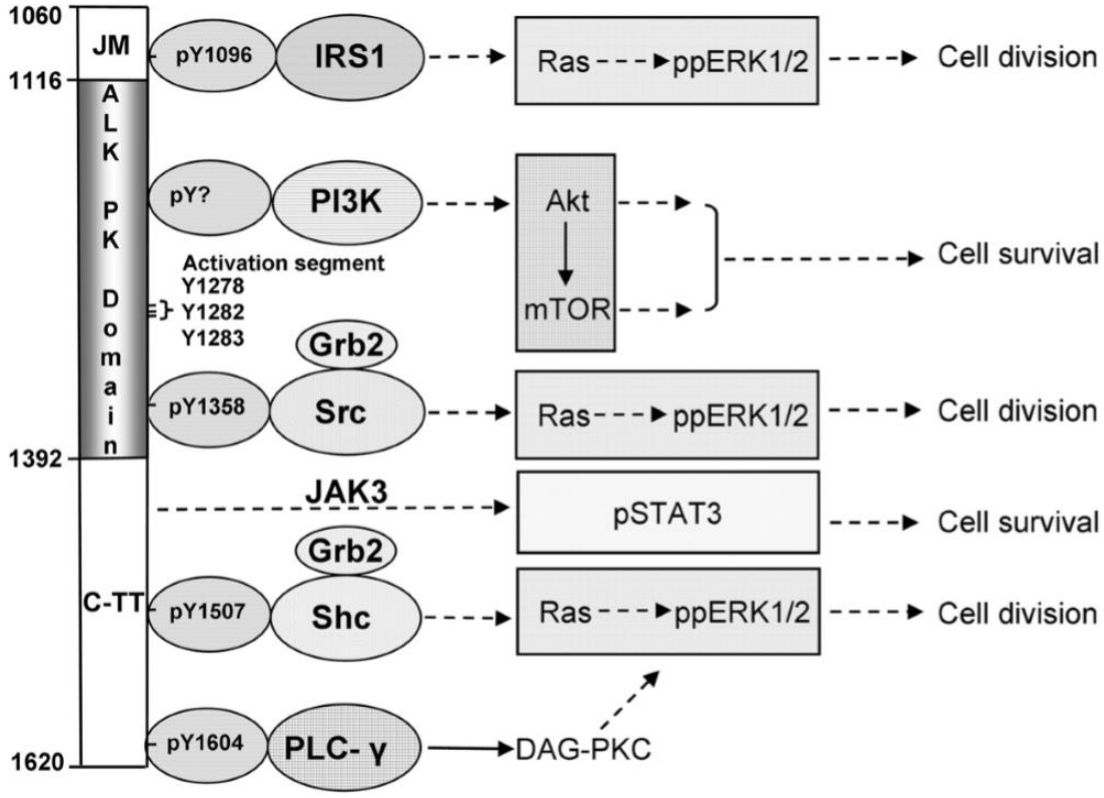


Figure 3 - Signaling pathways activated by ALK-fusion proteins. (Adapted from Roskoski, 2013)

BRAF

BRAF is a serine/threonine kinase that is part of the mitogen activated protein kinase (MAPK) signaling pathway (Peyssonnaux & Eychène, 2001). This pathway is involved in many processes related to cell proliferation, growth and survival (Sánchez-Torres, Viteri, Molina, & Rosell, 2013). Like its isoforms ARAF and CRAF, BRAF belongs to the Rapidly Accelerated Fibrosarcoma (RAF) family of kinases. In healthy tissue, the MAPK pathway is activated as a result of growth factors binding to a trans membrane receptor tyrosine kinase (Seger & Krebs, 1995). This results in a cascade of protein activation in which BRAF is activated by RAS after which it in turn phosphorylates MEK. This cascade

results in the activation of ERK which is then transferred to the nucleus where it activates transcription factors that promote cell proliferation and inhibit apoptosis (Figure 4)(Sánchez-Torres et al., 2013). BRAF as a driver in cancer is most commonly seen in melanoma where ~66% of tumors show activating BRAF mutations (Davies et al., 2002). In NSCLC, the percentage of cases showing activating mutations of BRAF is significantly lower at ~3% of patients (Brose et al., 2002). While BRAF mutations are a relatively rare occurrence in NSCLC, due to the large amount of total cases, BRAF remains an important drug target in this condition. The most common activating mutations of BRAF are a substitution of the val600 residue with glutamic acid (V600E) or lysine (V600K) (Bradish & Cheng, 2014; Luk et al., 2015). V600 mutation results in the formation of constituently active BRAF monomers (Spagnolo, Ghiorzo, & Queirolo, 2014). For treatment of NSCLC expressing V600 mutated BRAF, the inhibitors dabrafenib and vemurafenib are used in clinical practice. Using these inhibitors, a better progression free survival is obtained compared to using conventional chemotherapy (Amaria et al., 2018; Larkin & Fisher, 2012).

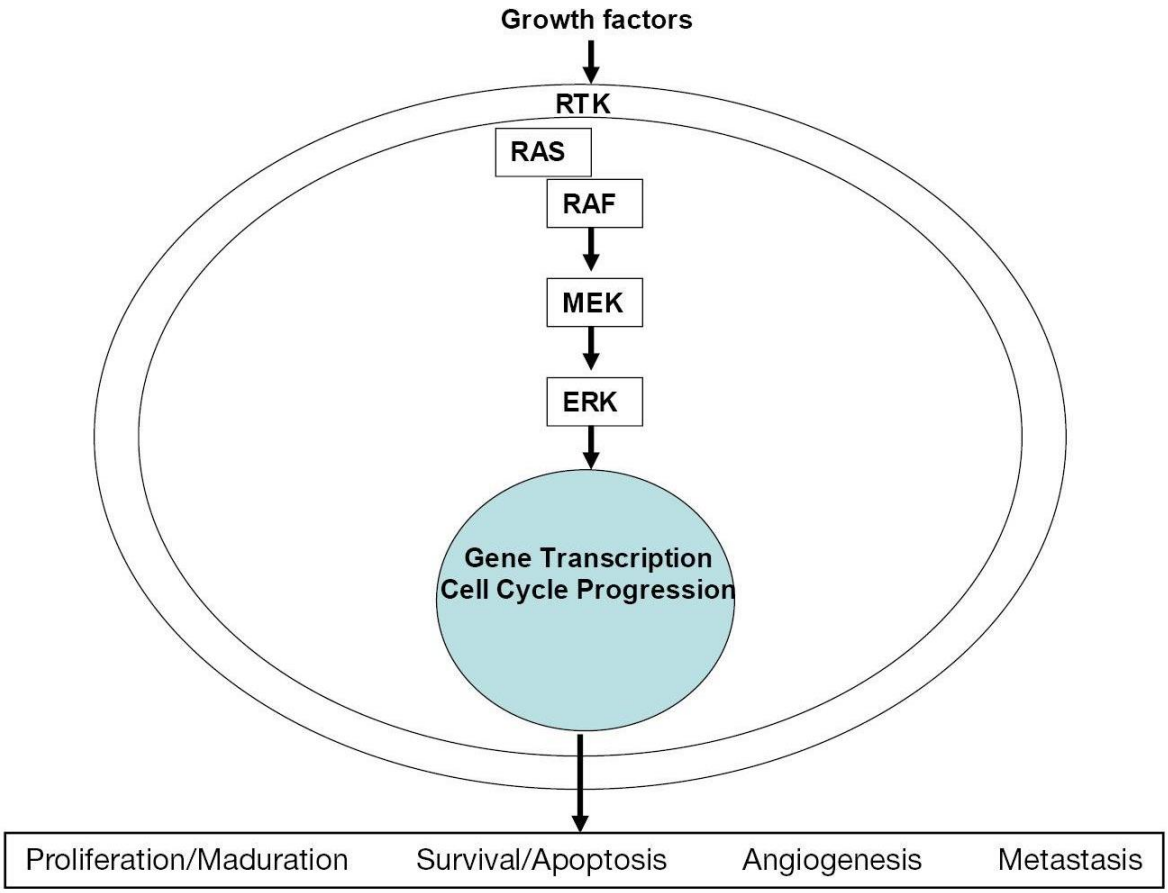


Figure 4 – Overview of the RAS-RAF-MEK-ERK signaling pathway. Oncogenic activation of BRAF can drive tumor growth through deregulation of this pathway (adapted from Sánchez-Torres et al., 2013)

Drug resistance

Based on results from clinical trials, it has been shown that treatment of kinase dependent NSCLC with kinase inhibitors is very effective. However, while these therapies are highly effective initially, treatment with any of the current inhibitors leads to the emergence of drug resistance. This is a major limiter on the efficacy of these treatments (Bean et al., 2008; Ryohei Katayama, 2018; R. J. Sullivan & Flaherty, 2013). A common mechanism by which drug resistance emerges is through

secondary mutations of the kinase (Bean et al., 2008; Li et al., 2017).While such a mutation might render he current therapy ineffective, it is in many cases possible to switch to another inhibitors that is still effective (J. C.-H. Yang et al., 2017). However, while the drug sensitivity of some common mutations is well known, in cases of uncommon or novel mutations medical professionals often have little to no information available to help inform clinical decision-making.

In order to resolve this issue, tumor boards at the UMCG have implemented computational methods such as homology modeling and molecular docking to predict the effects of mutations on drug efficacy. In this manner, drug sensitivity has been assessed in over 70 cases in which little to no literature information was available. This has allowed for more informed clinical decision making in cases with novel resistance mutations (van Kempen et al., 2018). While this method can provide medical experts with the information they need to make informed clinical decisions, interpreting the data currently requires a large time investment by knowledgeable experts. Molecular docking can lead to the generation of up to hundreds of potential binding poses per drug which currently have to be evaluated manually. This is a major limitation on the scalability of such methods which prevents them from being implemented more broadly in the treatment of patients.

In order to resolve this limitation, molecular docking filters were designed to identify binding poses with predictive value in regard to the efficacy of kinase inhibitors. The filters that were designed for this are based around well understood geometric restrictions on the various interactions between the drugs and the kinases in the bound form. This report describes the setup of a computational pipeline starting from the mutation to the filtering of the docking dataset. The filters will be validated by examining the filtering results for mutations with known sensitivity or resistance to kinase inhibitors. This way, the report looks to show if these filters can be used to greatly reduce the work required to extract predictions of drug efficacy from computational data. Should that be the case, then this potentially enables a broader implementation of computational methods to predict drug efficacy in patients expressing novel kinase mutations. This could allow for patients to be treated in a more personalized and effective manner, while preventing side-effects from ineffective treatments.

Materials and methods

Homology modeling

The first step to computationally predicting the effects of mutations on the efficacy of drugs is to determine the effects of said mutation on the structure of the target protein. An accessible method for doing so is homology modeling. Using this method, the structure of a protein of interest can be predicted based on the known structure of a protein with high sequence identity (Waterhouse et al., 2018). In order for reliable models to be generated, the sequence identity of the template structure with the structure of interest should be as high as possible (Gromiha, Nagarajan, & Selvaraj, 2019). In the case of the kinases ALK, BRAF and EGFR, this criterion can be met by making use of publicly available crystal structures for these kinases from the RCSB protein databank (Berman, 2000). By using these crystal structures as templates, the difference between the generated models and the template structures is limited to only a few amino acids. Since this is only a minor difference, it is expected to result in the generation of reliable high quality homology models.

In order to predict drug efficacy accurately, an effort has to be made to ensure that the homology models generated can accurately represent the protein as it interacts with the inhibitors. For this purpose, a selection of PDB crystal structures was made based on four selection criteria. The first criterion was related to the presence of mutations in the template structures. Generally, this should be avoided as it likely reduces the sequence identity with the mutant structures. For this reason, in ALK and EGFR only wild-type crystal structures were selected as templates. For BRAF, structures containing the V600E mutation were also used. This was done since this is the most common primary BRAF mutation seen in patients, and the inhibitors are specifically indicated for cases with V600 mutated BRAF (Holderfield, Deuker, McCormick, & McMahon, 2014). Because of this, many of the secondary mutations where it would be useful in practice to predict BRAF inhibitor efficacy concerns mutations that include an alteration of the V600 residue. In order to ensure the highest possible sequence identity in BRAF, mutations containing V600 mutations will be modeled using templates containing the V600E mutant. Furthermore, some of the BRAF crystal structures available through the PDB contain several surface residue mutations to assist in expression, since these are not believed to impact the overall structure of the kinase, these will be ignored for the purpose of model selection (Tsai et al., 2008).

The second selection criterion is related to the structure of the kinase domain. The inhibitors discussed in this report bind their respective kinases in specific conformations. In the computational setup used, it is therefore important to select templates that show conformations compatible with inhibitor binding. To ensure this, two aspects of the protein structure will be used as inclusion criteria for template structures. These aspects are the positioning of the regulatory αC-helix as well as the DFG motif (Figure 5). These aspects are indicative of the catalytic activity of the kinase conformation (Roskoski, 2013; Vijayan et al., 2015). In bioactive conformations, both the αC-helix and the DFG motif are in the so called “in” position. The difference between the αC-in and αC-out position can be seen in Figure 6. The αC-in position allows the helix to form the interactions necessary for catalytic activity (Roskoski, 2017). The DFG-motif refers to the orientation of the “Asp-Phe-Gly (DFG)” residues in the activation loop. The DFG motif is referred to as “out” if it is rotated away from its conformation in the active state (Treiber & Shah, 2013). As can be seen in Figure 7C/D, the state of the DFG motifs has a large impact on the positioning of the activation segment of the kinase and determines the accessibility of an allosteric binding pocket (Nagar et al., 2002). The EGFR and ALK inhibitors referred to in this report can bind to catalytically active conformations of their respective kinase (Roskoski, 2016). In order to properly assess binding in the computational setup, templates will therefore be selected to have the αC-in and DFG-in conformation. For the BRAF inhibitors dabrafenib and vemurafenib however, templates will be selected that are DFG-in but αC-out. This is because the sulfonamide group in these drugs causes a shift of the αC-helix to the out position, making them incompatible with rigid αC-in structures (Roskoski, 2016).

Thirdly, a requirement was set for the resolution of the structure. Structures with high resolution are not useful for the purpose of predicting drug binding since there is much uncertainty in the location of the atoms in the structure. Because of this, structures with a resolution above 3 Å were not used for the modeling.

The last selection criterion is used to ensure that the static structures selected are in a conformation that is compatible with the binding of inhibitors. The binding of ligands to proteins can cause changes in their structure which is commonly referred to as “induced fit” (Morando et al., 2016). When

working with static models, this can greatly impact the models ability to properly bind the inhibitors (Grebner et al., 2016). Because of this, only crystal structures co-crystallized with one of the inhibitors that will be used for docking were selected for the modeling step.

Since the homology models are generated based on models that have an inhibitor co-crystallized, it is important to take into account that the induced fit effect can cause these models to show bias towards binding the co-crystallized drug. To highlight the effect co-crystallization with specific ligands has on the obtained crystal structure, Figure 8 shows the variation in the location of residues in EGFR crystal structures bound to various inhibitors. As can be seen in this figure, significant variation in the location of residues can be seen in many of the residues surrounding the binding pocket. To mitigate this problem, ensemble docking will be used. This entails that every drug is docked on a set of homology models based on different crystal structures (Amaro et al., 2018). This way, a larger portion of the conformational space of the kinase can be explored to more accurately identify which mutations hinder drug binding.

To find models that suit the template criteria, a PDB search was done for EGFR, BRAF and ALK structures in complex with the inhibitors mentioned previously. Whether or not these models passed the template criteria was assessed through visual inspection in PyMOL. The results of this analysis can be seen in Table 1,

Table 2 and Table 3. Based on the template criteria, the selection of templates was as follows: For EGFR, the structures 1M17, 4G5J, 4WKQ, 4I23 and 4ZAU met all criteria and were selected as templates. In BRAF, 3OG7, 4XV2, 4RZV and 5CSW were selected. Though the 4RZV model does contain an additional R509H mutation, this model was still selected as the mutation is well known to prevent BRAF dimerization but has not shown an effect on inhibitor binding (Poulikakos et al., 2011).

After model selection, the public modeling server SWISS-MODEL was used for the generation of homology models. This is a well-established protein modeling server based on the ProMod3 modeling engine (Waterhouse et al., 2018).

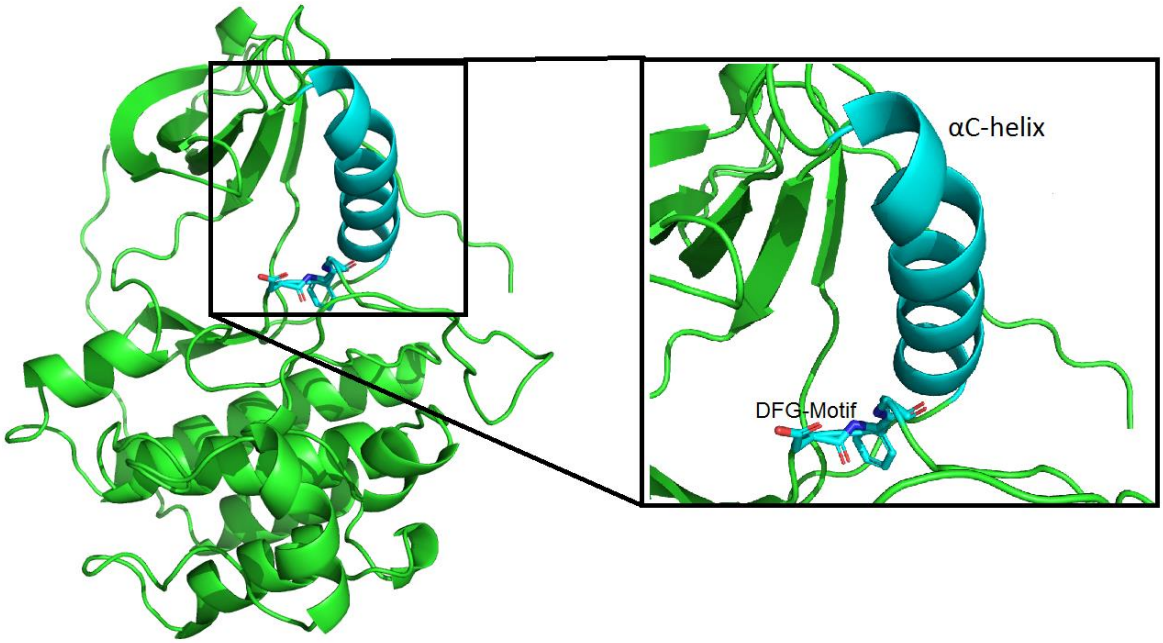


Figure 5 – EGFR kinase domain and ATP binding pocket. Highlighted are the regulatory αC-helix as well as the DFG-motif. These aspects are used in the selection of template crystal structures for the generation of homology models.

Table 1 – EGFR crystal structures taken from the PDB. Structures were assessed on being in the active conformation, the presence of existing mutations and their resolution.

EGFR models	Ligand	Mutation	Resolution	αC & DFG in
4HJO	erlotinib	V924R	2.75Å	No
1M17	erlotinib	-	2.6Å	Yes
4I22	gefitinib	V924R/L858R/T790M	1.71Å	No
4WKQ	gefitinib	-	1.85Å	Yes
3UG2	gefitinib	G719S/T790M	2.5Å	Yes
2ITO	gefitinib	G719S	3.25Å	Yes
2ITY	gefitinib	-	3.42Å	Yes
2ITZ	gefitinib	L858R	2.8Å	Yes
4G5J	afatinib	-	2.8Å	Yes
4G5P	afatinib	T790M	3.17Å	No

4I23	dacomitinib	-	2.8Å	Yes
4I24	dacomitinib	T790M	1.8Å	No
4ZAU	Osimertinib	-	2.8Å	Yes

Table 2 - BRAF crystal structures taken from the PDB. Structures were assessed on being in the active conformation, the presence of existing mutations and their resolution.

BRAF models	Ligand	Mutation	Resolution	αC-helic	DFG-motif
5CSW	dabrafenib	-	2.66Å	out	in
4XV2	dabrafenib	V600E	2.5Å	out	in
5HIE	dabrafenib	β3-αC del	3Å	malformed	in
4RZV	vemurafenib	R509H	2.99Å	out	in
3OG7	vemurafenib	V600E	2.45Å	out	in

Table 3 - ALK crystal structures taken from the PDB. Structures were assessed on being in the active conformation, the presence of existing mutations and their resolution.

ALK models	Ligand	Mutation	Resolution	αC & DFG in
2XP2	crizotinib	-	1.9Å	Yes
2YFX	crizotinib	L1196M	1.7Å	Yes
4ANQ	crizotinib	G1269A	1.76Å	Yes
4ANS	crizotinib	L1196M/G1269A	1.85Å	Yes
5AAA	crizotinib	L1198F	1.73Å	Yes
5AAB	crizotinib	C1156Y/L1198F	2.2Å	Yes
5AAC	crizotinib	C1156Y	1.7Å	Yes
4MKC	ceritinib	S1281G	2.01Å	Yes
6MX8	brigatinib	-	1.96Å	Yes
5FTO	entrectinib	-	2.22Å	Yes
4CLI	lorlatinib	-	2.05Å	Yes
4CLJ	lorlatinib	L1196M	1.66Å	Yes

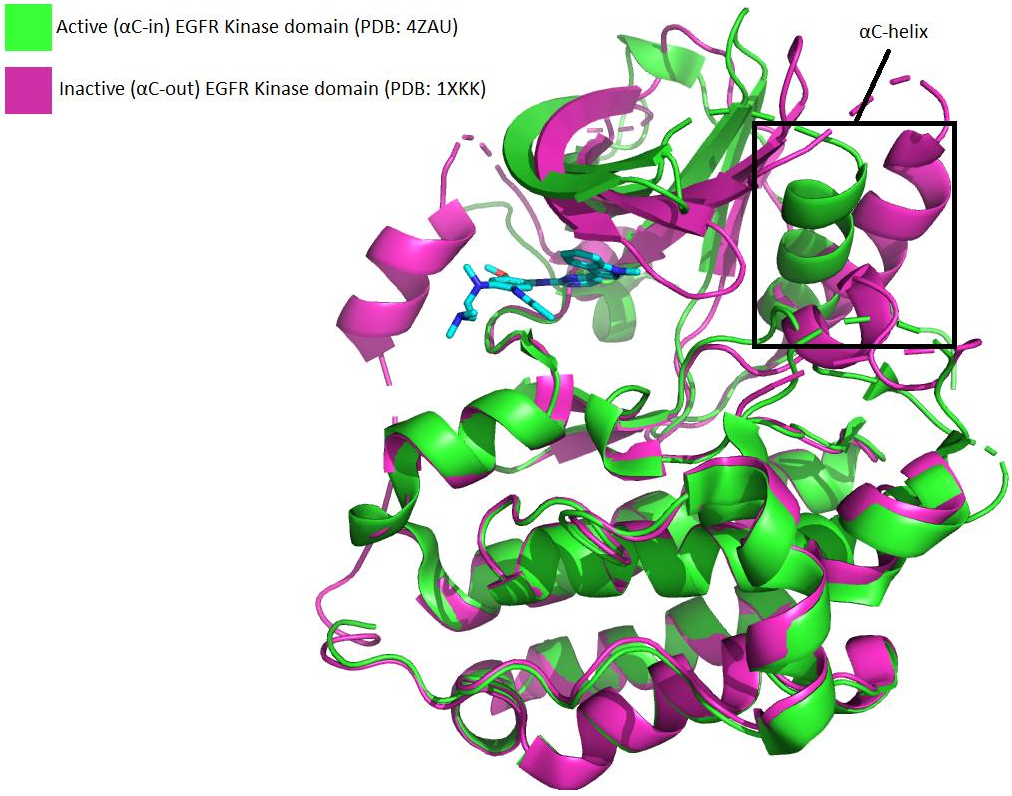


Figure 6 – Overlay of an active and inactive EGFR kinase domain. The αC-helix can be seen to be in the αC-in and αC-out conformations n the active and inactive conformations respectively.

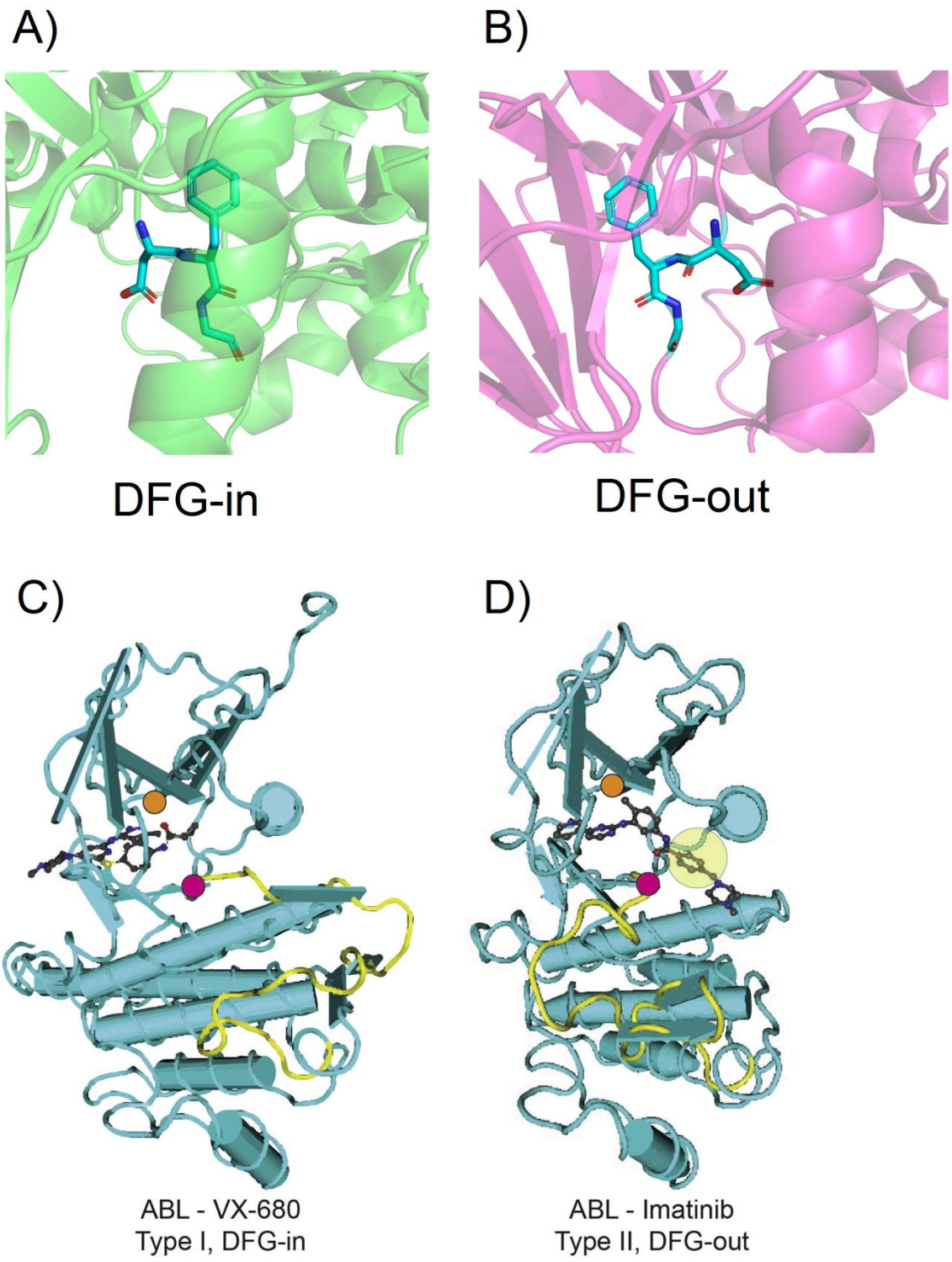


Figure 7 – The DFG motif and its effect on kinase structure. A) The DFG motif in the active "in"-conformation B) The DFG motif flipped in the inactive "out"-conformation. C) Kinase structure in the DFG-in conformation, the activation segment is highlighted in yellow. D) Kinase structure in the DFG-out conformation. The activation segment is highlighted in yellow and an allosteric pocket that is accessible in this conformation is highlighted in the yellow circle. (adapted from Treiber & Shah, 2013**)**

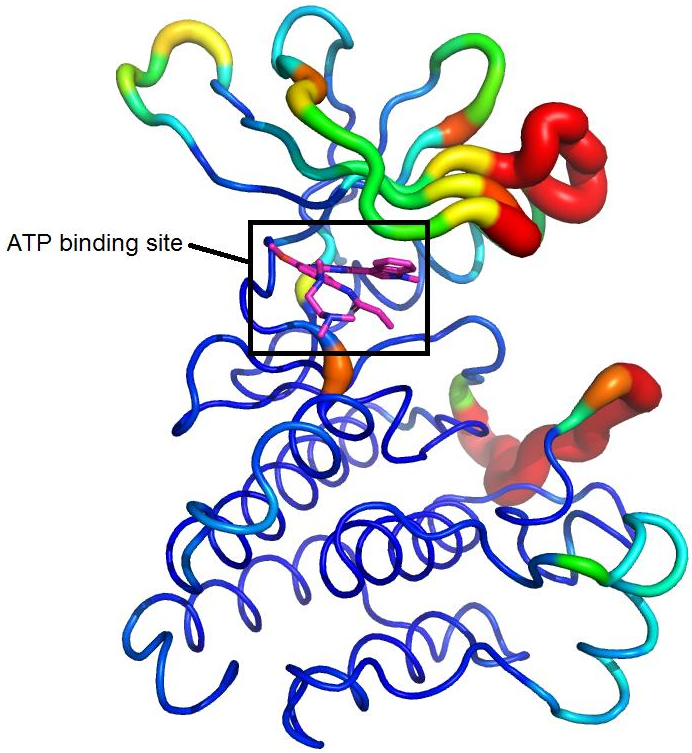


Figure 8 - Averaged location of alpha-carbon in the PDB structures used for homology modeling. The width of the line and the color indicate the root mean square deviation (RMSD) in the location of the atoms. Atoms with more variation in their location are represented by a greater line width. The RMSD values displayed range from 0.08 (dark blue), to >1.5 (red). (The displayed range of values was clamped at a max of 1.5Å to make differences in variability across the structure more visible) (PDB IDs used: 1M17, 2ITW, 2ITX, 2ITY, 2J6M, 4G5J, 4I23, 4WKQ, 4ZAU, 6D8E)

Molecular docking

After homology modeling, clinically used small molecule inhibitors for the modeled kinase will be docked used Smina. This is an open-source molecular docking program based on Autodock Vina (Koes, Baumgartner, & Camacho, 2013).

The inhibitors used for the docking were Erlotinib, gefitinib, afatinib, dacomitinib and osimertinib for EGFR, Dabrafenib and Vemurafenib for BRAF and alectinib, brigatinib, ceritinib, crizotinib, entrectinib and lorlatinib for ALK. When generating binding poses, Smina makes use of a randomized monte-carlo search algorithm to find ways in which the ligand can bind the protein that reduce the free energy of the system (Trott & Olson, 2009). In order to guarantee that the ATP binding pocket is properly searched for possible binding modes, both the area for the search and the search depth need to be properly defined. To make the entirety of the binding pocket accessible for the search algorithm, the search location will be defined as the boxed position of the co-crystallized ligand of a PDB crystal structure expanded by 5 angstrom in all directions. The ligand locations selected come from PDB structures 6MX8, 3OG7 and 4ZAU for ALK, BRAF and EGFR respectively. To guarantee that this area covers the binding pocket in its entirety, All homology models were aligned to the crystal structure from which the ligands location was used to define the area. After the alignment, the search area was validated visually using PyMOL (Figure 9).

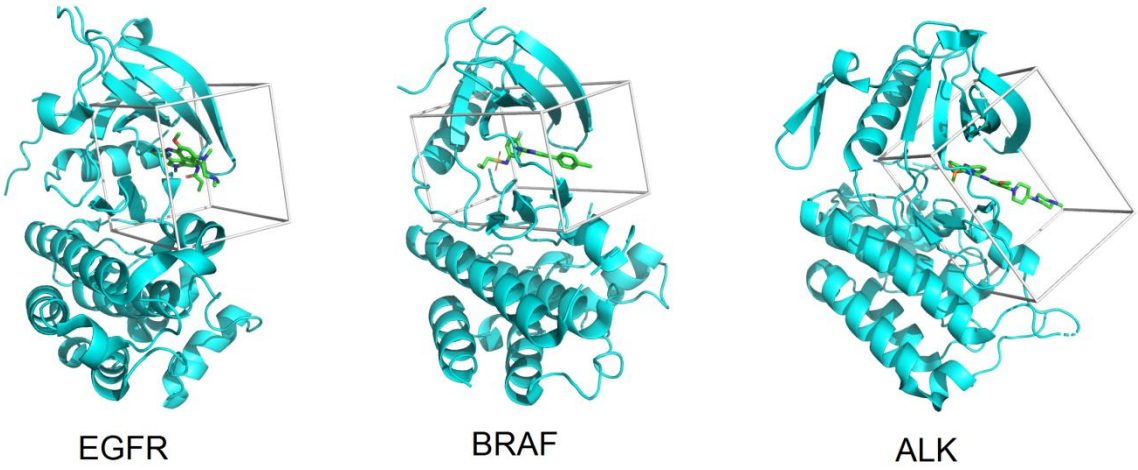


Figure 9 – Search areas defined for docking of inhibitors on structures of EGFR, BRAF and ALK. The boxes outlining the search area can be seen to cover the binding pockets entirely.

For the docking runs, the search exhaustiveness was set to 32 and SMINA was set to look for at most 25 poses within an energy range of 10kcal/mol of the highest predicted affinity pose found. Other settings were left at their default. Using these settings, Smina was used to obtain a list of up to 25 possible binding poses for each drug with a predicted binding affinity for each pose in kcal/mol. Since the search function used relies on randomization, docking of the ligands was repeated in triplicate to increase the reproducibility of the search.

In order to validate that binding poses closely resembling the drug binding modes seen in crystal structures can be generated in this manner, cross docking was performed. Here, the inhibitors for each kinase were docked onto the PDB crystal structures for the templates selected for the generation of homology models without introducing any mutations. To ensure the process is similar to what will be done with the homology models and to fill out missing residues, the PDB crystal structures were first submitted to SWISS-Model using their original FASTA sequences as the target. After static docking, the extent to which the generated poses resembled the known binding modes was quantified by calculating a root mean square deviation for each pose. When calculating the RMSD for these poses, it has to be noted that not all atoms in the drug interact with the kinase in the bound state. Parts of these drugs are exposed to the solvent, allowing for much variation in the position of these atoms. Because of this, using an all atom RMSD would not provide an adequate indicator for the quality of a generated binding pose. An example of this is shown in Figure 10, this figure shows a SMINA generated pose for gefitinib that maintains all interactions with the kinase that the crystal structure pose forms. However, since the solvent exposed tail doesn't align with the crystal structure binding pose, the all atom RMSD (2.83Å) doesn't reflect this. To rectify this, a selection of atoms was made to represent the part of the drug whose location is strictly defined in the bound state. The first selected point makes use of a characteristic hydrogen bond these inhibitors form with the hinge region of the kinase. An overlay of the EGFR, BRAF and ALK inhibitors in which this bond is highlighted can be seen in Figure 11. The second and third atom were selected to be the atoms furthest removed from the bond that are still in the same plane as the hydrogen bond acceptor atom. These three points define the location of a ring structure whose location is strictly determined by a combination of the hydrogen bond and hydrophobic interactions. Additional points were added based on other interactions formed between the drug and kinase. Many of the inhibitors

contain an aryl group that occupies a hydrophobic pocket near the ATP binding site. These include the EGFR inhibitors Erlotinib, Gefitinib, afatinib and dacomitinib, as well as both BRAF inhibitors and the ALK inhibitors crizotinib, ceritinib, brigatinib and entrectinib. Since the positioning of this group is well defined by hydrophobic interactions, a point was added to define the position of this ring. All atoms selected for the RMSD calculation are shown in Figure 12, Figure 13 and Figure 14.

During the re-docking, poses with an RSMD of 2Å or lower will be accepted as highly similar to the crystal structure binding pose. This cutoff was chosen as it is commonly used to identify similar binding poses (Ravindranath, Forli, Goodsell, Olson, & Sanner, 2015). Should such a binding pose be found, then it will be concluded that Smina is able to generate binding poses similar to the crystal structure binding mode for the drug in question.

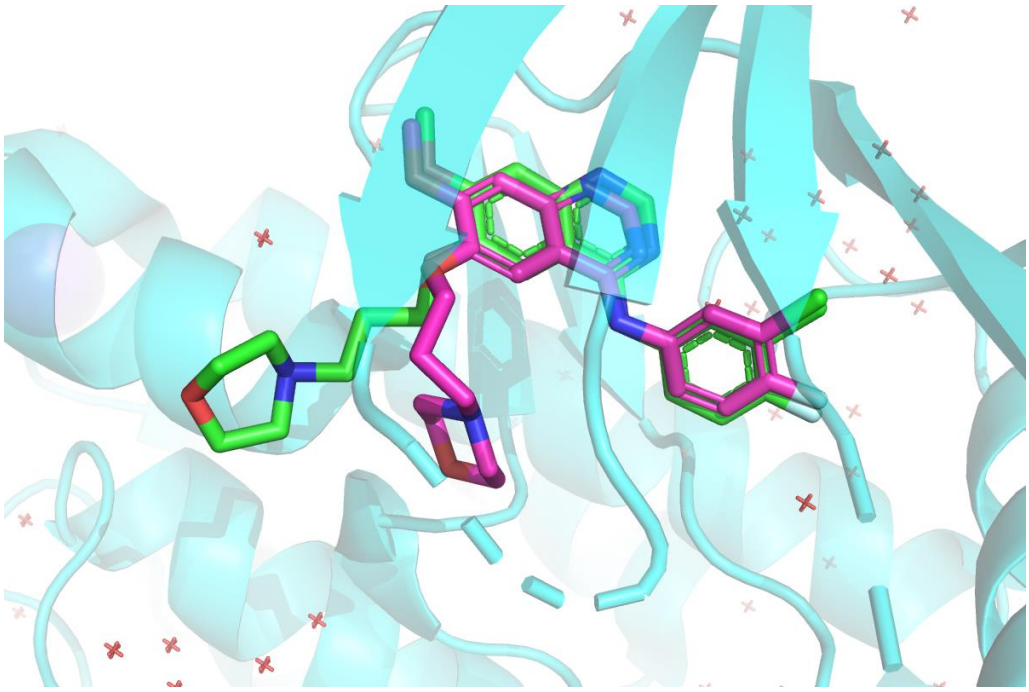


Figure 10 – SMINA generated gefitinib binding pose (magenta), superimposed on gefitinib crystal structure binding pose take from PDB accession code 4wkq. All atom RMSD between these poses is 2.83Å.

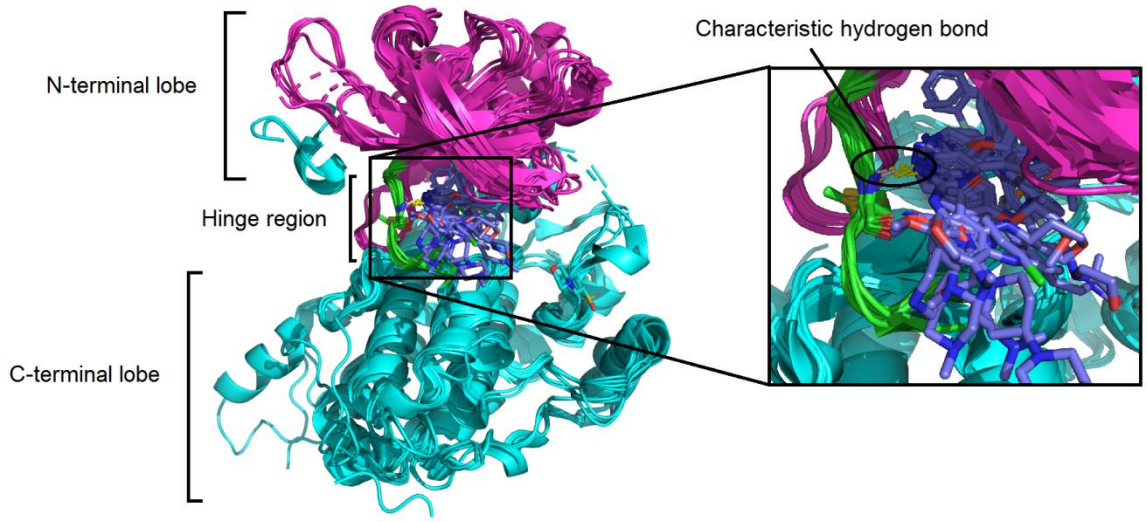


Figure 11 - Overlay of PDB crystal structures of EGFF, ALK and BRAF kinase domains in complex with their inhibitors. A characteristic hydrogen bond formed with the hinge region of the kinase domain is highlighted. PDB accession codes of structures shown are 1M17, 4G5J, 4WKQ, 4I23, 4ZAU, 3OG7, 4XV2, 2XP2, 5J7H, 4CLI and 5FTO

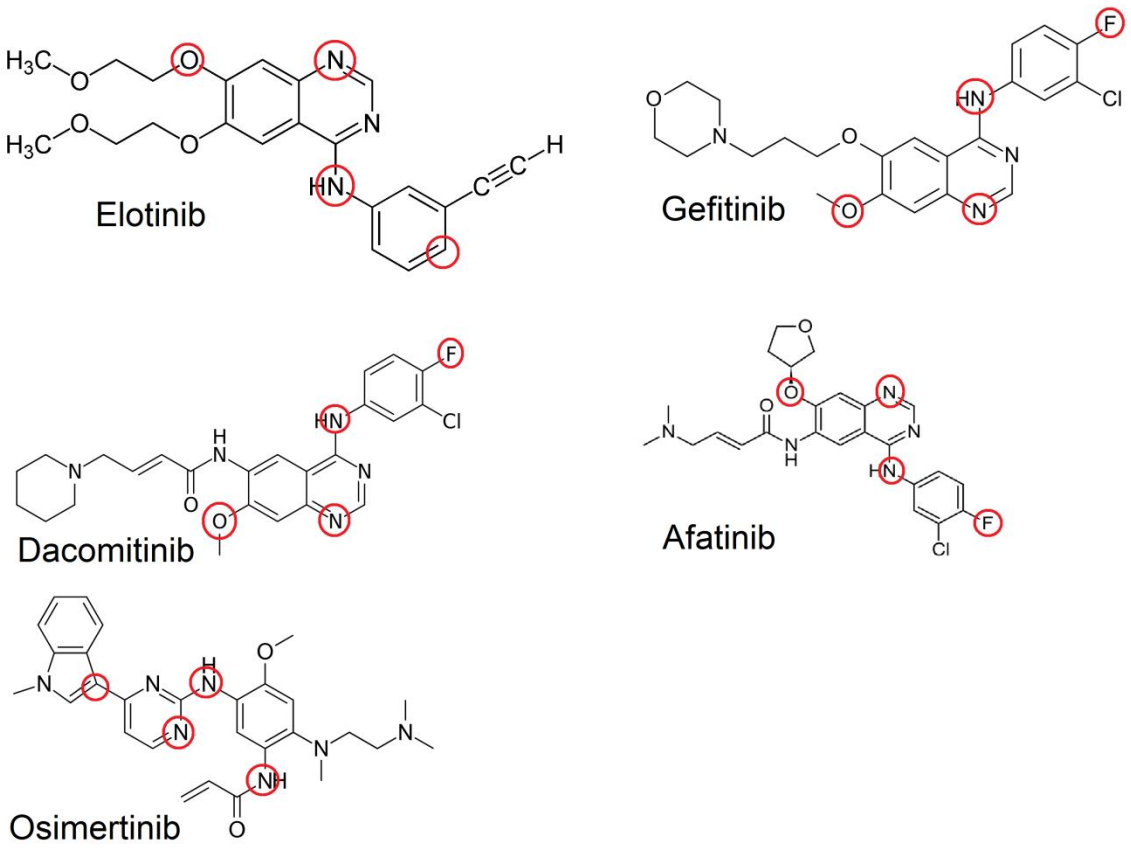


Figure 12 – EGFR inhibitors used during docking. Highlighted in red circles are the atoms used to calculate the RMSD value for each pose relative to a crystal structure pose.

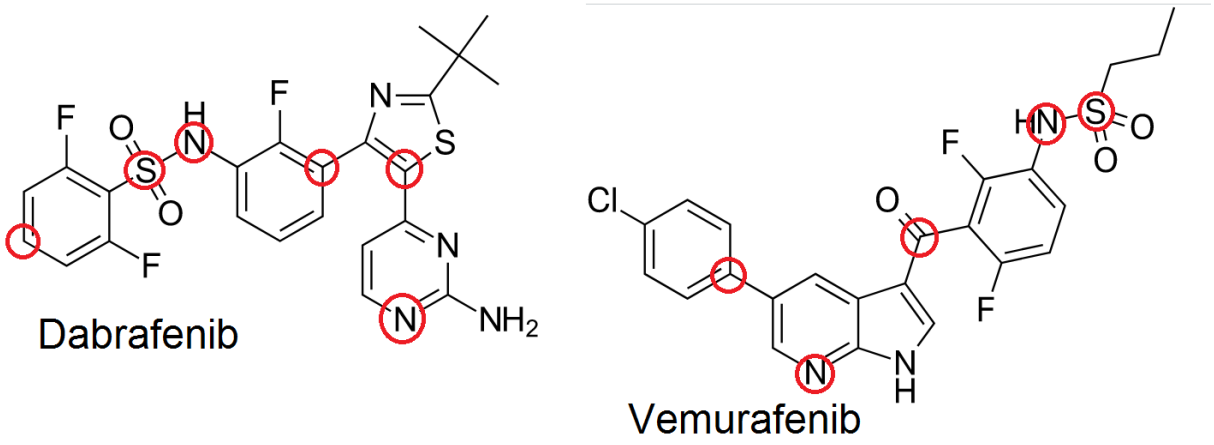


Figure 13 - BRAF inhibitors used during docking. Highlighted in red circles are the atoms used to calculate the RMSD value for each pose relative to a crystal structure pose.

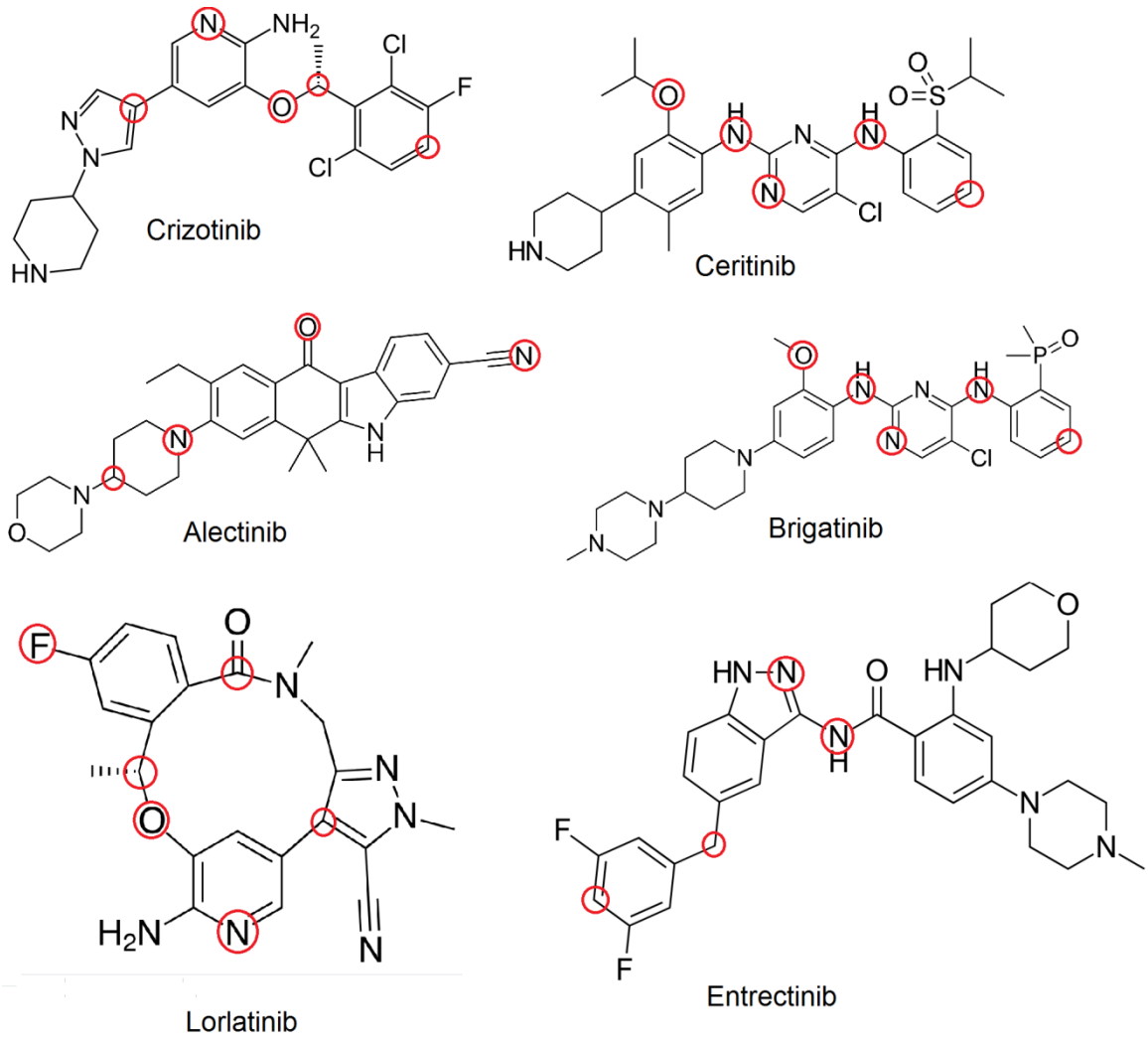


Figure 14 - ALK inhibitors used during docking. Highlighted in red circles are the atoms used to calculate the RMSD value for each pose relative to a crystal structure pose.

Design of docking filters

The manner in which ALK, BRAF and EGFR inhibitors interact with the kinase protein is known from crystal structures. However, many of the poses generated by Smina do not match this binding mode. These poses often lack interactions necessary for proper binding such as the characteristic hydrogen bond with the kinase hinge region. Since these binding poses do not resemble the known binding mode, the data on binding affinity provided by these poses is not clinically relevant. In order to reduce the amount of data that needs to be reviewed, these poses can therefore be filtered out.

For this purpose, a total of three filters will be designed to identify poses similar to the known binding mode from the docking dataset. The first filter will be based around the RMSD value as described previously. This filter can be used to identify poses that retain hydrophobic interactions and defines the correct orientation for the drug in the binding pocket. The second filter that will be designed is based around the strength of the hydrogen bond with the hinge region of the protein. A normalized measure of the strength of this bonds can be calculated based on its geometric features (Irbäck, Mitternacht, & Mohanty, 2009). Aspects of this that determine the strength of a hydrogen bond are the length of the bond and the angles between the donor, hydrogen and acceptor atoms (R. Kumar, Schmidt, & Skinner, 2007). The effect of distance on hydrogen bond strength can be approximated using the Lennard-Jones potential (Jones, 1924). This potential can be expressed as:

$$V_{LJ} = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$$

Here, ε is the depth of the potential well. Since we are working with normalized measures, this variable has a value of 1. σ is the finite distance at which the potential equals 0. In hydrogen bonds this is ~1.5 Å (Dannenberg, 1998). The variable r is the distance between the hydrogen bond and the acceptor atom in angstrom.

The angle component of bond strength can be calculated using a formula from a paper by Irbäck et al. (Irbäck et al., 2009):

$$v(\alpha,\beta) = \begin{cases} (\cos\alpha\cos\beta)^{1/2} & if\ \alpha,\beta > 90° \\ 0 & Otherwise \end{cases}$$

The variables in this formula are as follows: α is the angle between the donor, hydrogen and acceptor, β is the angle between the donor hydrogen, the acceptor atom and the averaged position of all atoms forming covalent bonds with the acceptor atom. This averaged position is used since its geometry dictates that it is in a direct line with the acceptor atom and the average position of electrons involved in the hydrogen bond. Figure 15 displays a 2D representation of the angles used. The normalized bond strength is expressed as the product of the Lennard-Jones potential and the angle component.

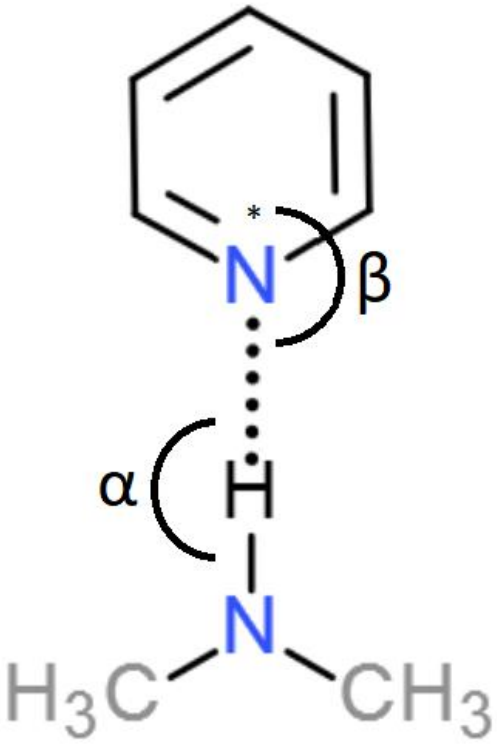


Figure 15 - 2D view of an NHN hydrogen bond. The * symbol shows the averaged positions of the carbons connected to the acceptor atom. α and β are the angles used in the bond strength calculations.

To determine how strong the hydrogen bond needs to be for it to be comparable to those found in crystal structures, the strength of this bond will be determined in PDB crystal structures of kinase inhibitors in complex with EGFR, BRAF and ALK. The strength of the hydrogen bonds in these structures will be used as a reference point for determining the minimum bond strength a pose should have to pass the filter.

The final position based filter will check for specific interactions for individual drugs. The first interaction that will be checked for is the correct placement of the ring structure in the hydrophobic pockets for the inhibitors that explore this pocket. This will be done by calculating the distance of a specified atom in this group relative to its expected location based on crystal structures. Should this distance exceed 5 angstrom, it can be assumed that the ring has been rotated outside of the hydrophobic pocket, causing the pose to fail the filter. Atoms used for all drugs this filter is applied to can be seen in Figure 16. Another drug specific filter will be implemented to indirectly determine the ability of poses for drugs that form a covalent bond with the kinase to form this bond. Since covalent bonds are not considered during the docking process, no direct indicator of covalent bonding is available in the docking data. Since the warhead groups that form the covalent bond do not otherwise form strong interactions with the kinase, their positioning will not be a reliable indicator for covalent bond formation. What can be done instead is to measure the distance between the pose and crystal structure locations of the atom connecting the warhead to the area of the drug whose location is defined in the RMSD filter. Should this atom have shifted by more than the van der Waals radius of a carbon, it will be assumed that the warhead will have shifted too far to still be in range for covalent bond formation. The atoms used for this can be seen in Figure 17.

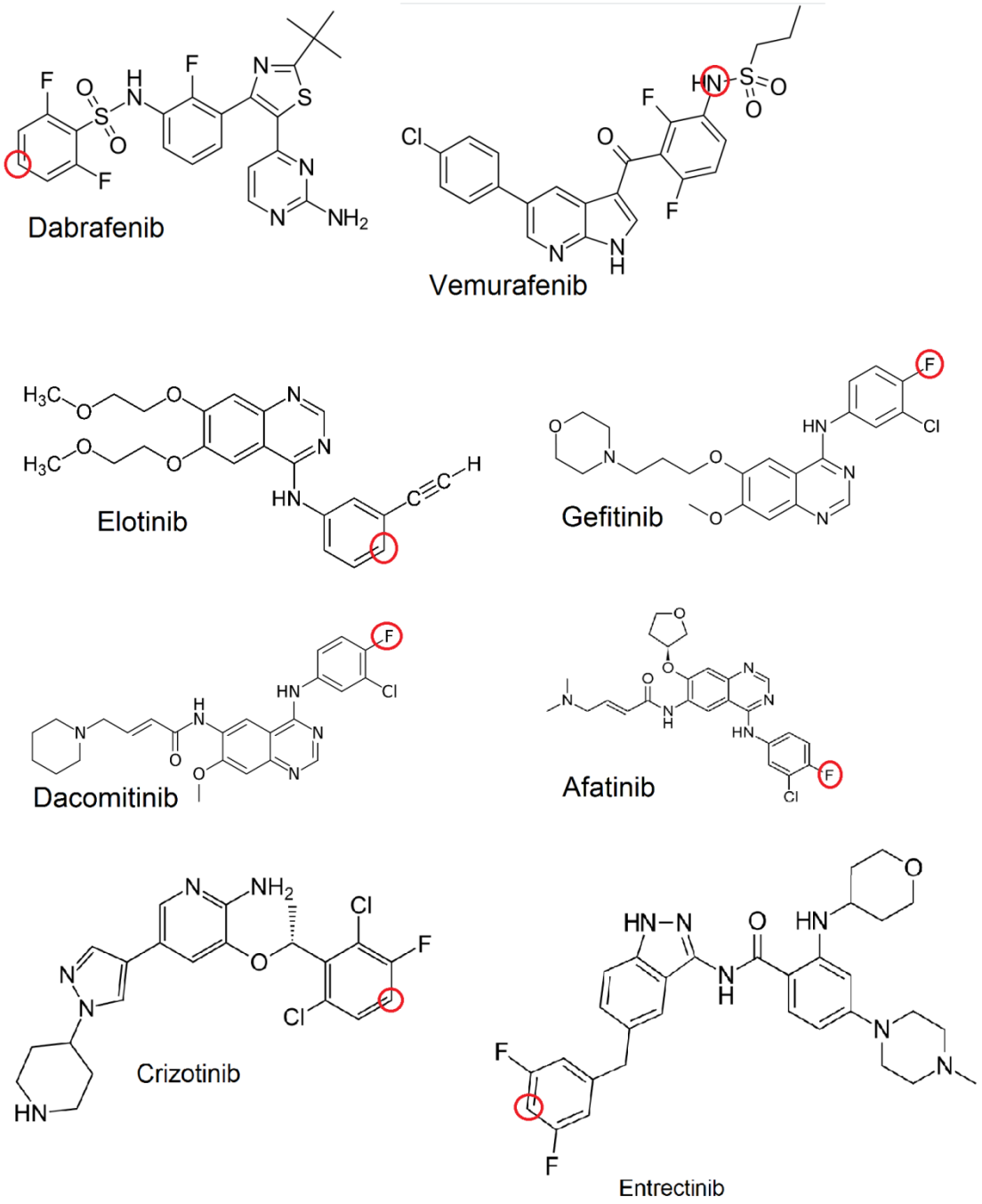


Figure 16 – Atoms used in the drug specific filters used to detect the aryl group occupying the hydrophobic pocket being flipped outside the pocket.

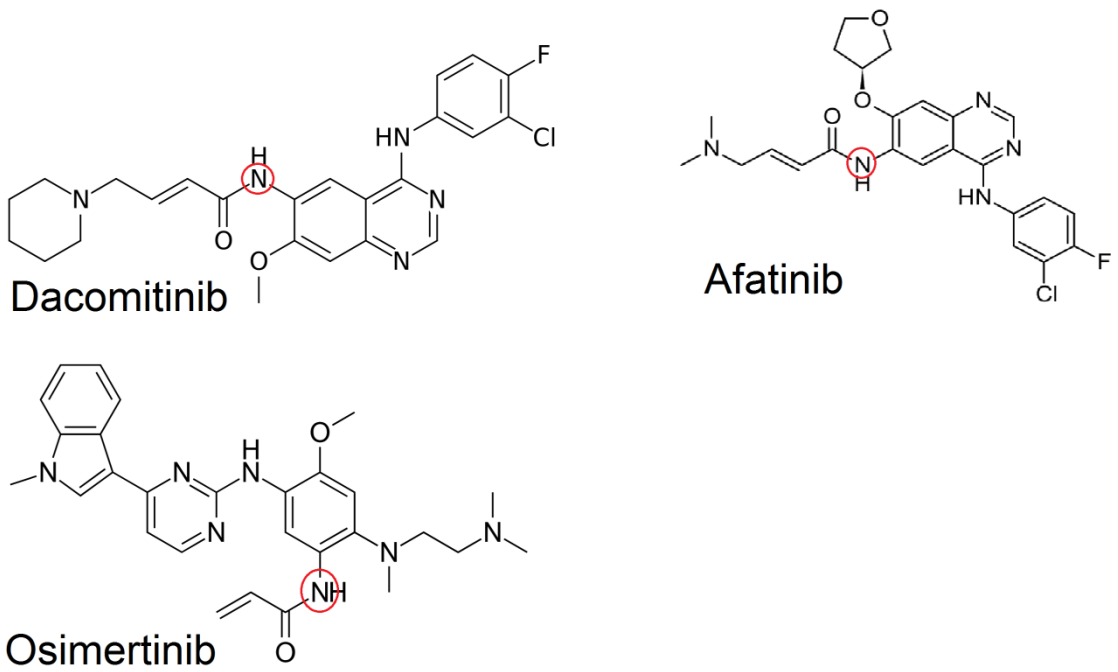


Figure 17 – EGFR inhibitors that form a covalent bond with the kinase. Highlighted in red are the atoms that will be used to assess if a generated pose is likely to be able to form this bond. Should a shift in the location of this atom be greater than the van der waals radius of a carbon, the pose fails the filter.

Aside from the filters looking at structural aspects of the binding pose, a separate filter will be designed that looks at the predicted binding affinity as provided by SMINA. In the cross-docking dataset, the highest affinity pose passing all filters for each drug will be marked as a reference for how strong the interaction is predicted to be in sensitive mutations. Should the binding affinity of a generated pose be significantly lower than this value, it will be assumed that it is no longer able to effectively compete with ATP binding. As an initial cutoff for this filter, an increase in binding affinity of over 10% will be used as the maximum increase to filter out weak binding poses.

Validation of docking filters

In order to validate if the docking filters can be used to detect poses useful in predicting drug resistance, they will be applied to docking data for a set of mutations known to be sensitive or resistant to specific inhibitors. An initial set of mutations was obtained from case studies. This list was further supplemented with the most commonly reported mutations in the COSMIC mutation database for which information about drug sensitivity was available. The mutations that were selected for EGFR, ALK and BRAF as well as the resistance they are known to confer can be seen in Table 4, Table 5 and Table 6. For each of the selected mutations, homology models were generated with SWISS model for each template crystal structure. Static docking was then run on these models as described earlier. After docking, each pose in the generated dataset was assigned a score based on the number of filters it passes. After scoring the poses individually, a score for each drug on individual models was assigned based on the largest number of filters passed in all docking poses for that drug. The scores across different models for the same mutant were then assessed for the presence of patterns indicating whether or not the drug is capable of binding the mutant kinase. Should models for mutations that are known to be resistant to specific inhibitors consistently show no poses passing all filters, then this will be used to show that the filters are able to both select

binding poses that are consistent with the known binding modes from crystal structures as well as that the data can be used as an indicator for the resistance conferred by various drugs.

Table 4 – EGFR mutation in NSCLC with known impact on kinase inhibitor efficacy.

EGFR mutation	Known sensitive	Known resistant	Reference
G719A/C/S	All	-	(Y. Kobayashi et al., 2015)
T790M	Osimertinib	Erlotinib, gefitinib, dacomitinib, afatinib	(S. Kobayashi et al., 2005; S. Wang, Cang, & Liu, 2016)
T790M/G796R/S	-	All	(S.-H. I. Ou et al., 2017)
T790M/L718Q	-	All	(Callegari et al., 2018)
T790M/L792F	-	All	(Y. Kobayashi et al., 2017)
T790M/L792H	-	All	(Q. Zhang et al., 2018)
T790M/C797S	-	All	(Ercan et al., 2015)
G724S	-	Osimertinib	(Fassunke et al., 2018; Oztan et al., 2017)
T854A	Erlotinib, gefitinib	Dacomitinib, Afatinib, Osimertinib	(Bean et al., 2008; Goyal, Jamal, Shanker, & Grover, 2015)
E764_S752del/ T790M/ P794L	Afatinib	Osimertinib	(van Kempen et al., 2018)
A763insFQEA	Erlotinib		(Voon, Tsui, Rosenfeld, & Chin, 2013)
D761Y		Erlotinib, gefitinib	(Balak et al., 2006)
E709A	All	-	(Huang, 2004; Y. Kobayashi et al., 2015)
E709K	All	-	(Y. Kobayashi et al., 2015)
E746_A750del	All	-	(Fukuoka et al., 2011)
L747P		Erlotinib, gefitinib	(G. Yu et al., 2015)
L747S		Erlotinib, gefitinib	(Chiba et al., 2017)
L747_S752del	All	-	(Fukuoka et al., 2011)
L858R	All	-	(Rosell et al., 2012)
L858R/D761Y	Erlotinib, gefitinib	-	(Balak et al., 2006)
L858R/ V843I	-	Erlotinib, gefitinib, dacomitinib, afatinib	(Stewart, Tan, Liu, & Tsao, 2015)
L861Q	All	-	(Lynch et al., 2004; J. C. H. Yang et al., 2015)
S768I	Dacomitinib, afatinib	Osimertinib	(Nasu et al., 2018; J. C. H. Yang et al., 2015)
V769insASV	Erlotinib, gefitinib	-	(Vyse & Huang, 2019)

Table 5 - ALK mutations in NSCLC with known impact on kinase inhibitor efficacy.

ALK mutation	Known sensitive	Known resistant	Reference
D1203N	-	Crizotinib	(Rolfo et al., 2014)
E1210K/D1203N	-	Crizotinib, brigatinib	(Heuckmann et al., 2011; Yoda et al., 2018)
G1202R	Lorlatinib	Crizotinib, ceritinib, brigatinib, alectinib	(Akamine, Toyokawa, Tagawa, & Seto, 2018; Hatcher et al., 2015)
G1269A	Ceritinib	Crizotinib	(Friboulet et al., 2014)
L1196M	Ceritinib, brigatinib, entrectinib, lorlatinib	Crizotinib	(Ryohei Katayama, 2018)
L1198F	Crizotinib	Lorlatinib	(Li et al., 2017)
V1180L	Brigatinib, ceritinib, lorlatinib	Crizotinib, alectinib	(R. Katayama et al., 2014; Lin et al., 2018; Takigawa, 2018)
C1156Y/L1198F	Crizotinib	Lorlatinib	(Shaw et al., 2016)
S1206C	Crizotinib	Ceritinib	(Umapathy, Mendoza-Garcia, Hallberg, & Palmer, 2019)
E1210K/S1206C	-	Crizotinib, brigatinib	(Gainor et al., 2016)
I1171N/S/T	Ceritinib, brigatinib, lorlatinib	Crizotinib, alectinib	(Sehgal et al., 2019)
E1210K	-	Crizotinib	(Gainor et al., 2016)
C1156Y	Alectinib, entrectinib, lorlatinib	Crizotinib, ceritinib	(Friboulet et al., 2014; Shaw et al., 2016; Song, Wang, & Zhang, 2015; I. Sullivan & Planchard, 2016)
F1174C	Alectinib	Crizotinib, ceritinib	(Carneiro et al., 2018; Friboulet et al., 2014)
F1174L	Ceritinib	Crizotinib	(Debruyne et al., 2016)
F1174V	Alectinib	Crizotinib	(S. H. Ou et al., 2016)
L1152R	Alectinib, brigatinib	Crizotinib, ceritinib	(Sabari et al., 2017; Tchekmedyian, Ali, Miller, & Haura, 2016)
1151Tins	Alectinib	Crizotinib, ceritinib	(Friboulet et al., 2014; Tran & Klempner, 2016)

Table 6 - BRAF mutation in NSCLC with known impact on kinase inhibitor efficacy.

BRAF mutation	Known sensitive	Known resistant	Reference
V600D/E/G/K/M/R	All	-	(Chapman et al., 2011; Falchook et al., 2012)
V600E/ T529I/M/N	-	Dabrafenib, vemurafenib	(Whittaker et al., 2010)
V600K/L505H	-	Vemurafenib	(Hoogstraat et al., 2015)
G466V	-	Dabrafenib, vemurafenib	(Wan et al., 2004)
V600E/L514V	-	Vemurafenib	(J. Wang et al., 2018)
G469A	-	Vemurafenib	(Porcelli et al., 2015)
G469L	-	Vemurafenib	(Gautschi et al., 2013)
L597Q/R/S/V	Vemurafenib	-	(Dahlman et al., 2012)
V600E/D587N	All	-	UMCG case
V600E/H608Y	All	-	UMCG case
V600K/I617V	All	-	UMCG case
K601E	Dabrafenib	-	(Murali, Menzies, & Long, 2012)

E. Coli expression of ALK

In order to allow for experimental validation of the predictions made using the computational methods, a method is needed that would allow for fast and reliable production of kinase proteins. This would allow for the use of methods like microscale thermophoresis (MST) or protein kinase assays to validate the predicted binding of drugs as well as their ability to inhibit mutant kinase function (Jerabek-Willemsen et al., 2014). A first step towards such a method was made by performing an expression screening, testing the ability of different *E. coli* cell lines to express the human ALK tyrosine kinase domain.

For this purpose, cDNA encoding for the wild-type human ALK tyrosine kinase domain was ordered from Eurofins Scientific. The DNA sequence was digested using *NcoI* and *HindIII* and ligated into pETM-11 and pETM-30 vectors for screening. These vectors were subsequently used to transform *E. coli* cell lines BL21 Star (DE3), Origami 2(DE3), Rosetta 2(DE3)pLysS and LOBSTR. Cells were transformed using the Heat Shock method and plated overnight at 37^{o}C on LB-Agar medium with the correct antibiotics (Table 7). A single colony was then selected and transferred into 10mL LB agar with the correct antibiotics for the strain and vector. Following overnight incubation at 37^{o}C, the LB was added to a 1L volume of LB and incubated at 37^{o}C on a 180RPM shaker. When the culture reached an optical density (OD) of 0.6 at 600nm, 1mM IPTG was added to induce protein expression overnight.

After overnight expression, cells were spun down for 30 minutes at 5000rpm after which the supernatant was discarded. The pellet was then dissolved in buffer containing 20.8mM $NaHPO_4$, 19.2mM NaH_2PO_4, 0.5M NaCl, 5% v/v glycerol and 10mM imidazole. Lysozyme was added to the buffer followed by incubation at 4deg for 30 minutes. The solution was then sonicated for a total pulse time of 2 minutes with 0.5 sec pulse time per cycle and 3s intervals.

After sonication, solution was centrifuged for 1 hour at 18000rpm at 4 degrees. The pellet and supernatant were separated and the supernatant was used in nickel column purification. During purifications, a total of 5 wash steps were performed using 20mL of 20.8mM $NaHPO_4$, 19.2mM NaH_2PO_4, 0.5M NaCl, 5% v/v glycerol buffer. This was followed by 5 elution steps using the same buffer with 300mM imidazole added. Fractions collected during purification as well as a small amount of resuspended pellet were run on an SDS-PAGE gel to look for the presence of and assess the expression of ALK in each combination of cell line and vector.

Table 7 – Antibiotics used after transforming each cell line with pETM-11 and pETM-30 vectors. Kanamycin is included in all cell lines since the resistance is conferred by the vector.

***E. coli* strain**	**Antibiotics**
BL21 Star (DE3),	50 µg/mL kanamycin 34 µg/mL Chloramphenicol
Origami 2(DE3)	50 µg/mL kanamycin 12.5 µ/mL Tetracyclin
Rosetta 2(DE3)pLysS	50 µg/mL kanamycin 34 µg/mL Chloramphenicol
LOBSTR	50 µg/mL kanamycin

Results and discussion

Cross docking

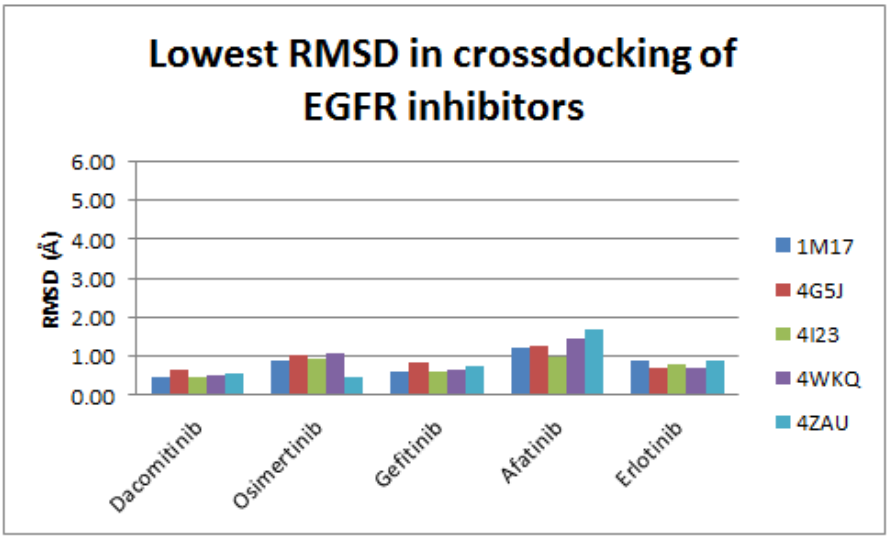
Before docking was performed on mutated homology models, cross docking using the inhibitors for EGFR, BRAF and ALK was performed on the crystal structures used as templates in the modeling step. The purpose of this was to validate that, when docking on these models, SMINA is able to generate binding poses for the drug that closely resemble known crystal structure binding poses. In order to ensure the docking procedure is the same as it is for the homology models, the crystal structures were first run through SWISS-Model with their original FASTA sequence as the target. This way, missing residues will be filled out in the same manner as will be the case in the homology models. SMINA was then used to dock the inhibitors on the structures of their respective kinase.

After the docking was completed, RMSD values were generated for each pose using the locations of the atoms noted in the materials and methods section. To indicate how similar poses in the docking dataset are to crystal structure poses, the lowest RMSD poses for each drug on the models of its respective kinase were used. The RMSD values for these poses can be seen in Figure 18. In this figure, it can be seen that for all drugs except for ceritinib, the lowest RMSD pose found never exceeded 2Å. This indicates that SMINA is able to generate binding poses similar to what is seen in crystal structures. For ceritinib however, all binding poses generated on the 5FTO template exceeded 5Å. This shows that binding poses similar to the crystal structure binding mode for this drug will likely not be generated on this model regardless of the effects additional mutations may have on drug binding. Because of this, docking results on models using this template cannot be considered reliable for ceritinib. When interpreting docking results for this drug, results for 5FTO templates should therefore not be considered.

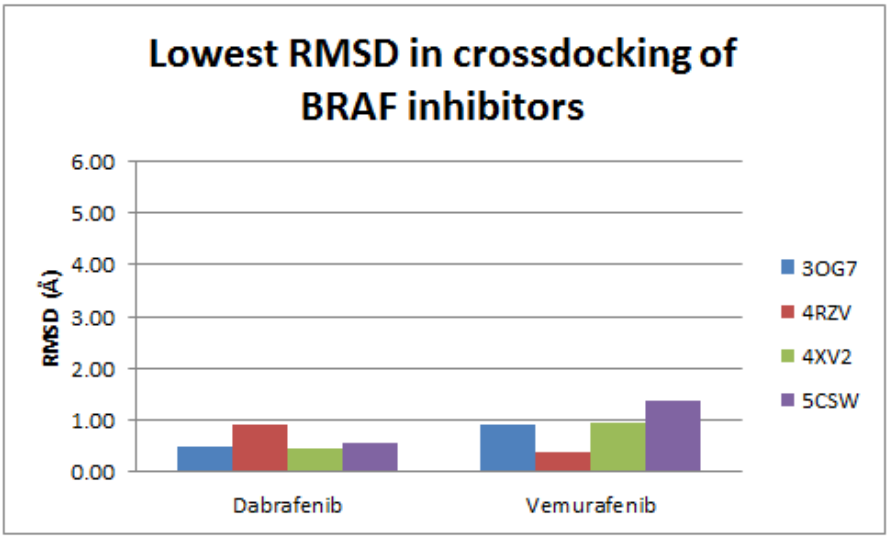
To verify the similarity of the low RMSD poses with the crystal structure binding modes, a visual inspection of generated poses was done in PyMOL. An example of a pose considered to closely resemble the crystal structure binding pose for an EGFR, BRAF and ALK inhibitor can be seen in Figure 19. Representative examples of poses for each drug can be found in supplementary figures: Figure S 1, Figure S 2 and Figure S 3. In the visual assessment, it was confirmed that the EGFR and BRAF inhibitors docked properly on all models. For ALK however, relevant differences were found between the crystal structure binding modes and the docking poses for alectinib, brigatinib, ceritinib and crizotinib. Firstly, the length of the characteristic hydrogen bond in generated poses for alectinib on the 2XP2, 4CLI and 5FTO templates exceeded 3Å in all poses. Since this bond is required for proper inhibitor binding, these templates lack a generates pose that accurately reflects how alectinib interacts with ALK. The same result was found for brigatinib and ceritinib on the 4CLI and 5FTO models. An example of the effect observed can be seen in Figure 20. Finally, it was found that crizotinib poses generated on the 4CLI template don't match the orientation of an aryl group that forms multiple hydrophobic interactions with ALK. In the generated poses, this group is flipped or rotated away from the pocket, breaking the hydrophobic interactions that serve to stabilize the drug inside the binding pocket (Figure 21). Since the generated ALK poses poorly resemble the known crystal structure binding modes, the predictions made about inhibitor efficacy with ALK mutations will likely reflect the known resistances very poorly.

Based on these results, binding of EGFR and BRAF inhibitors appears to be modeled properly. For ALK however, many models did not generate proper binding poses for multiple ALK inhibitors. Because of this, it can be expected that the predictions made about drug efficacy will better reflect the known resistances in EGFR and BRAF compared to ALK.

A



B



C

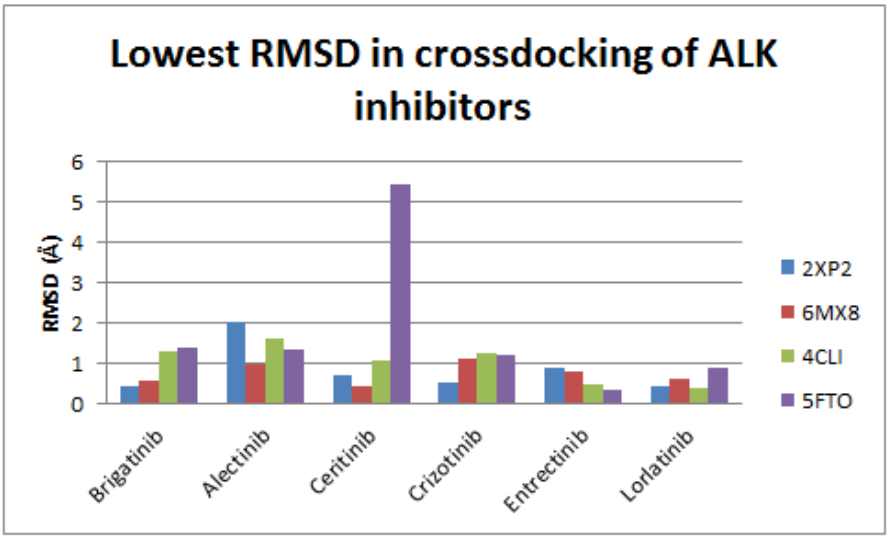


Figure 18 – Lowest RMSD for binding poses generated for each drug during cross-docking. A, B and C show the minimum RMSD calculated using the selected points outlined in the materials and methods section for EGFR, BRAF and ALK inhibitors respectively.

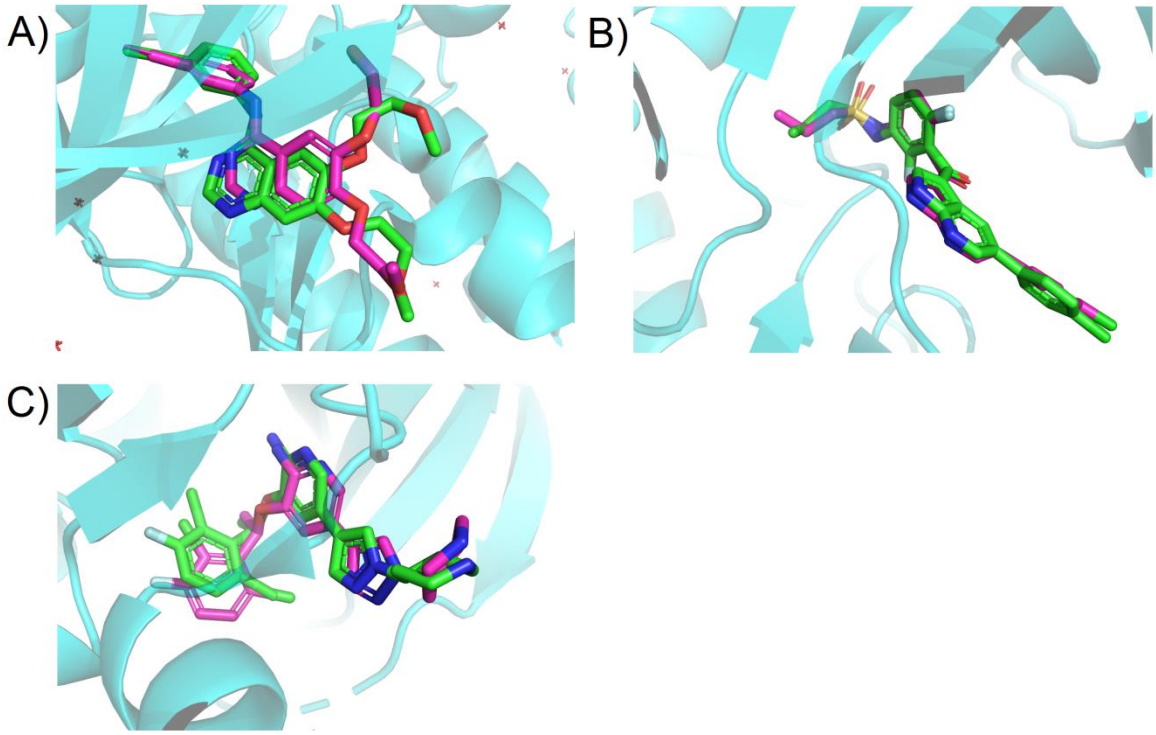


Figure 19 – Generated binding poses for Erlotinib (A), vemurafenib (B) and crizotinib (C) in complex with EGFR, BRAF and ALk respectively (shown in magenta). Crystal structure binding modes taken from crystal structures with PDB accession codes 1M17 (A), 3OG7 (B) and 3AOX (C) shown in green. Generated poses can be seen to closely resemble the crystal structure binding mode. Oxygen and Nitrogen atoms are colored red and blue respectively.

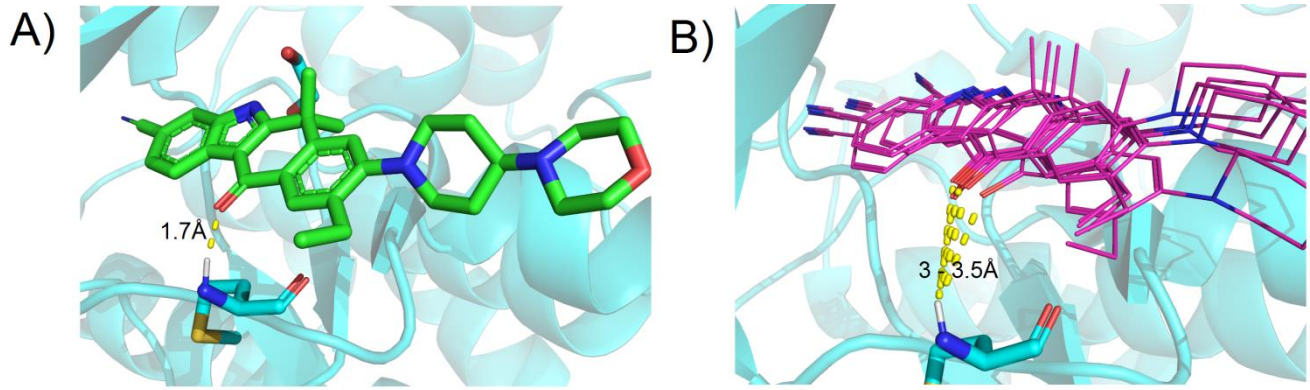


Figure 20 – A) Crystal structure binding mode for Alectinig (PDB accession code: 3AOX), a hydrogen bond with a bond length of 1.7Å can be seen. B) Docking poses for Alectinib representative of those generated on template structures 4CLI, 5FTO and 2XP2. While the overall orientation of the generated poses is similar to the crystal structure, the poses appear displaced causing a large increase in the hydrogen bond length, eliminating this interaction. Oxygen and Nitrogen atoms are colored red and blue respectively

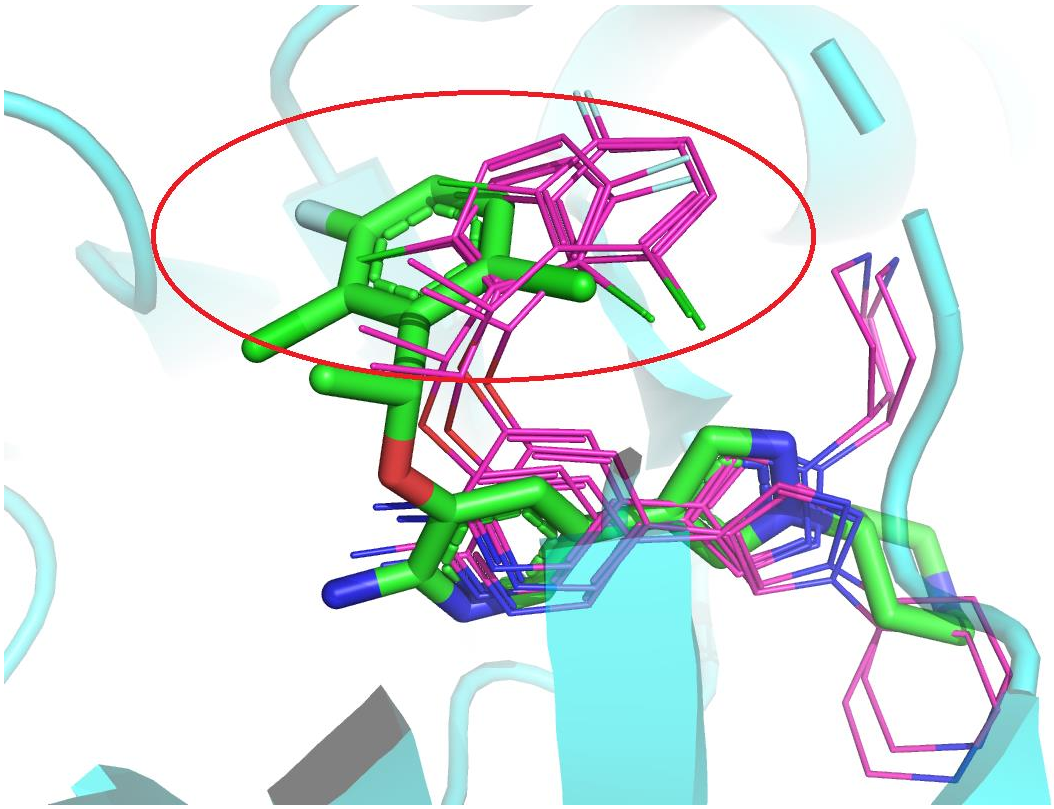


Figure 21 – Generated binding poses for crizotinib on the 4CLI template structure (shown in magenta) most closely resembling the crystal structure binding mode (shown in green). The aryl group highlighted in the red circle is flipped in the generated poses, causing interactions normally formed by this group to be lost. Oxygen, nitrogen, chlorine and fluorine atoms are colored red, blue, dark green and white respectively

Docking filters

A total of four docking filters were designed to help identify poses that can indicate that an inhibitor is capable of properly binding and thus inhibiting a mutant kinase. As described in the materials and methods section, the first filter removes poses with a predicted binding affinity more than 10% higher than predicted for the template structure. Another filter makes use of the RMSD of the selected atoms as described previously. A third filter checks for drug specific interactions such as the positioning of specific ring structures and the possibility for the formation of covalent bonds.

The final filter has to test for the presence of the hydrogen bond the inhibitors form with the hinge region of the kinase. In order to do this, a maximum strength this bond can have needs to be determined to serve as a cutoff value for filtering poses. To establish such a value, 52 crystal structures of ALK, BRAF and EGFR in complex with inhibitors were obtained from the PDB. The normalized bond strength of the characteristic hydrogen bond was determined in each structure as described in the materials and methods section. The normalized bond strengths can be found in the supplementary Table S 1. Within this set of structures, the maximum normalized hydrogen bond strength with the hinge region residue was found to be -0.1509. To account for the limited number of crystal structures available for determining a cutoff, half this value (-0.07545) was used as a maximum cutoff for hydrogen bond strength in the filter.

Filter validation

In order to assess the quality of the predictions made using the filtering setup, it was applied to the most common mutations with known clinical relevance for EGFR, BRAF and ALK.

EGFR

For EGFR, the mutations G719A/C/S, G724S, T790M, T790M/C797S, T790M/G796R, T790M/G796S, T790M/L718Q, T790M/L792F, T790M/L792H, T854A, A763insFQEA, D761Y, E709A, E709K, E746_A750del, L747P, L747S, L747_S752del, L858R, L858R/D761Y, L858R/V843I, L861Q, S768I and V769insASV were modeled. For all these mutations, information regarding their sensitivity to some or all available EGFR inhibitors is publicly available. Homology models for these mutations were generated using the crystal structures with PDB accession code 1m17, 4g5j, 4i23, 4wkq and 4zau as templates. The inhibitors erlotinib, gefitinib, afatinib, dacomitinib and osimertinib were docked on the generated homology models after which the docking poses were scored based on how many of the filters they pass. The results of this analysis are shown in Figure 22. The first point of interest consists of the filtering results for the crystal structure templates used to generate the other homology models. As can be seen in these results, all models pass all filters for each drug. This further confirms that docking on the selected models can produce poses highly similar to what is seen in crystal structures. Should changes to this be seen in the mutated models, this can be viewed as a potential indicator for drug resistance.

The mutations G719A, G719C and G719S are activating mutations located in the binding pocket. These can lead to the development of NSCLC by making it more energetically favorable for the kinase to adopt a catalytically active conformation (Y. Kobayashi et al., 2015). Since the template structures used for generating the homology models were already in the active conformation, it is expected that little effect will be seen on the docking results for activating mutations. These mutations do not confer drug resistance. In the docking results, it can be seen that the generated poses for the inhibitors pass all filters in the majority of models. The only exception to this is osimertinib. While these mutations do not confer resistance to osimertinib, the lack of passing poses for this drug is not entirely unexpected. Osimertinib is a drug specifically designed to be effective against T790M mutated EGFR (Soejima, Yasuda, & Hirano, 2017). The absence of this mutation leaves an area of unoccupied space directly next to the bound form of osimertinib. To illustrate this, Figure 23 shows and compares a binding poses generated for the G719A and T790M models of the 1M17 template. As can be seen, the docking for the non-T790M model results in the drug being moved further into the pocket to occupy the small cleft of unoccupied space. This causes these poses to fail the RMSD filter.

The T790M mutation is one of the most common causes of resistance against EGFR inhibitors (Gazdar, 2009). This mutation introduces a bulky methionine residue into a hydrophobic pocket that is explored by erlotinib, gefitinib, dacomitinib and afatinib. Figure 24 shows an overlay of crystal structure poses for these drugs onto the 4ZAU T790M model. As can be seen in this figure, the binding mode for these drugs is sterically hindered by the presence of the Met790 residue when it is in the orientation predicted by SWISS-Model. Due to the steric clash caused by this, T790M mutant EGFR is only sensitive to inhibition with osimertinib (H. Zhang, 2016). Within the set of mutations modeled, the T790M mutations as well as all other mutations containing T790M are expected to be resistant against erlotinib, gefitinib, afatinib and dacomitinib. These mutations are: T790M, T790M/C797S, T790M/G796R, T790M/G796S, T790M/L718Q, T790M/L792F and T790M/L792H. In the docking results for these mutations, the vast majority of models for these drugs failed to pass one or multiple filters. This provides a clear indications of drug resistance which aligns with the known resistance for these mutants. Furthermore, in the T790M single mutant results, osimertinib is

shown to pass all filters in all models. This indicates that osimertinib is effective in T790M mutant NSCLC. These results accurately reflect what is known from literature.

The T790M/C797S mutation is resistant against all clinically used EGFR inhibitors, including osimertinib. The mechanism for its resistance against osimertinib is the loss of a covalent bond that osimertinib normally forms with the cys797 residue. When looking at the docking data for this mutation, no indications of resistance against osimertinib can be seen. This is expected, since covalent interactions are not directly assessed during docking. While this could limit the predictive value of the filtering setup, since the covalent bonding EGFR inhibitors only form a single covalent bond all with this same cysteine residue, this shortcoming will not greatly affect the predictive value of the setup in novel mutations.

The mutations G796R, G796S and L718Q all interfere with osimertinib binding by abolishing critical hydrophobic interactions between EGFR and the inhibitor (Callegari et al., 2018; S.-H. I. Ou et al., 2017; Q. Zhang et al., 2018). When generating models for these mutations, the T790M mutation was added to the models as well. This was done since in cases described in literature, this mutation appeared after treatment with osimertinib on T790M mutated EGFR (Callegari et al., 2018; Q. Zhang et al., 2018). When looking at the docking results for these mutations, the results for T790M/G796R show a very clear indication of drug resistance. That this mutation would have a stronger effect on osimertinib binding compared to G796S is expected since an arginine residue is much larger than the serine residue. Besides eliminating a hydrophobic interaction, the G796R mutation also leads to steric hindrance with osimertinib binding, creating a very clear indication of resistance in the filtering data (Figure 25).

The G796S mutation is also known to hinders osimertinib binding (S.-H. I. Ou et al., 2017). Though the serine residue is considerably less bulky than the arginine introduced by the G796R mutation, it still interferes with the positioning of the osimertinib acrylamide group. Figure 26A and B show the clash with the crystal structure pose location of the acrylamide linker caused by the G796S mutation. In multiple models, this resulted in a shift of the acrylamide linker group away from the Cys797 residue. Figure 26C shows an example of a pose where such a shift was seen. This shift causes the pose to fail the filter checking for drug specific interactions and suggests that the G796S mutations could hinder the formation of the covalent bond by keeping the acrylamide warhead away from the cysteine residue. Since G796S fails to pass the filters in multiple models, this leads to the conclusion that this mutation could be resistance against osimertinib. This result is in line with previous literature detailing the resistance conferred by this mutation (S.-H. I. Ou et al., 2017).

The L718Q has a mechanism of resistance similar to G796R/S, in that it also eliminates a hydrophobic interaction between EGFR and osimertinib. Like was seen in those mutations, failure to pass all filters can again be seen in multiple models. As was the case with G796S, visual inspection of the generated poses revealed a shift in the acrylamide group greater than 1.7Å. This is expected to hinder covalent bond formation with the C797S residue. A representative example of a L718Q pose can be seen in Figure 27. Like G796R/S, L718Q is known to confer resistance to osimertinib, the docking results obtained therefore match existing literature (Callegari et al., 2018).

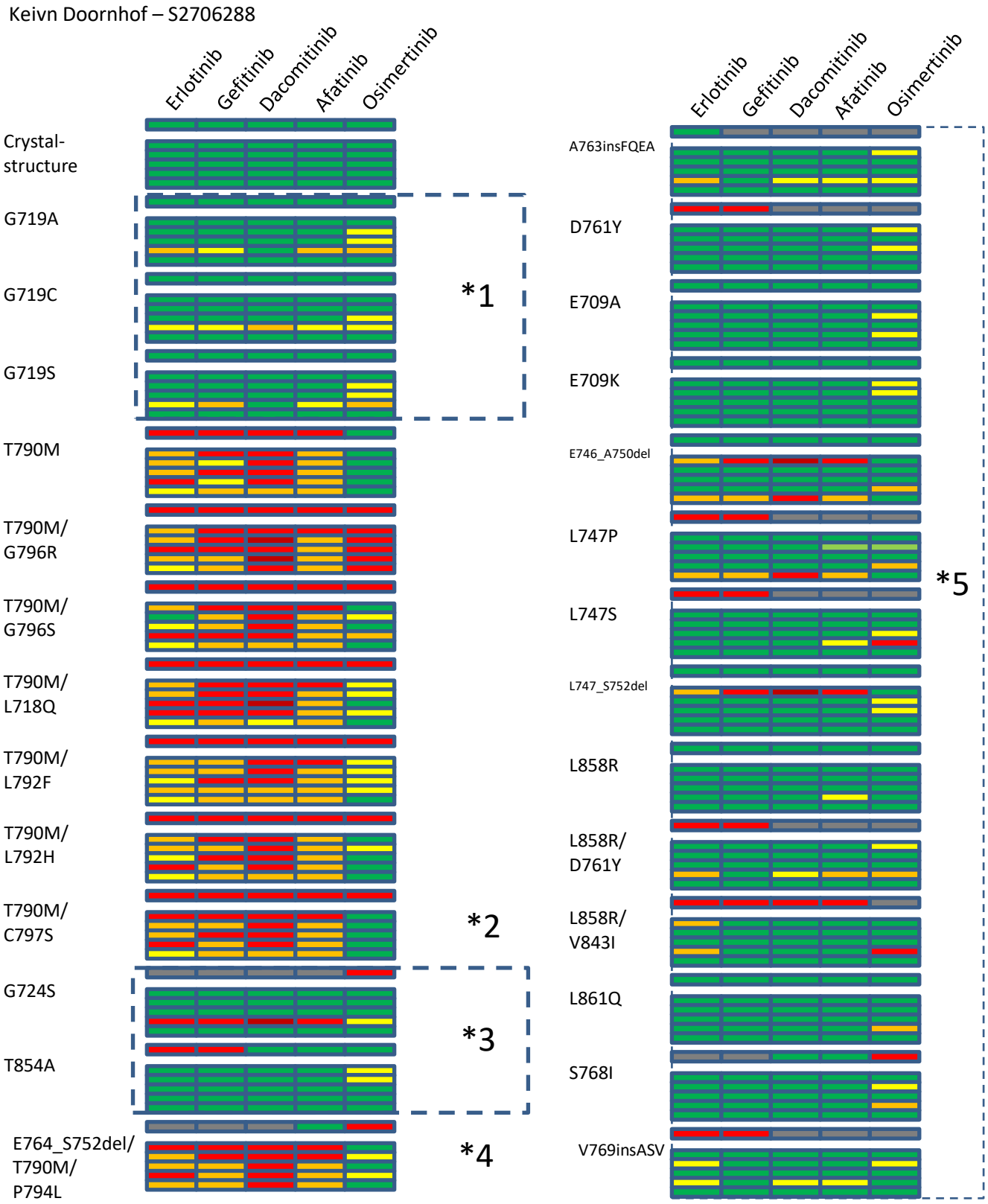
Mutations of the L792 residue are known to cause steric interference with the methoxy group of osimertinib (Y. Kobayashi et al., 2017; Q. Zhang et al., 2018). In the poses generated for the L792F models based on PDB structures 1m17, 4g5j, 4i23 and 4wkq, this resulted in a shift of the acrylamide

warhead that will likely prevent the formation of a covalent bond with Cys797. The only model in which the shift was not large enough to show in the filtering results is 4zau. Since the 4ZAU template structure was originally co-crystallized with osimertinib, induced fit provides a potential explanation for the pass seen in this model. In the case of L792H mutation, no clear indication of osimertinib resistance is seen in the filtering results. When inspecting the generated poses for osimertinib however, a displacement of the acrylamide group similar to what was seen in L792F could be seen. The displacement in this mutation was smaller than the van der waals radius of a carbon (1.7Å), which is why these poses did pass the filters. While a stricter filter could have more accurately predicted drug resistance in this case, doing so could also lead to false positives in detecting drug resistance since propensity for covalent bond formation is measured in a very indirect manner. This result shows that while static docking as used here can provide valuable information about drug resistance in many cases, more subtle changes in drug binding are likely to be overlooked.

The triple mutant case E746_S752del/P794L/T790M has been previously assessed at the UMCG tumorboard. Here this mutation was predicted to be resistant against osimertinib but sensitive to afatinib. The resistance and sensitivity to osimertinib and afatinib respectively was confirmed when the patient was treated using these drugs (van Kempen et al., 2018). When looking at the filtering data however, the results predict resistance against afatinib. A possible explanation for this is that during the analysis done at the tumorboard, a template model was used that contained the T790M mutation (PDB accession code: 4I24). In this case, this brings the starting sequence closer to the mutated target, which could increase the accuracy of the method. The use of template structures with fewer sequence differences might allow for more accurate predictions in this case.

The remaining resistance conferring mutations located in the ATP binding pocket are G724S, which is resistant against osimertinib, and T854A, which confers resistance to erlotinib and gefitinib. In both of these cases, the docking data aligns poorly with the known resistances conferred by these mutations. Visual inspection of the data similarly provided no indication of drug resistance. A potential explanation for this can be found in literature detailing potential mechanisms of drug resistance for the G724S and T854A mutations. Both of these mutations have an impact on the flexibility of the kinase structure. In both cases, it has been hypothesized that the resistance that is seen results from structural changes that are the result of this change in flexibility (Fassunke et al., 2018; Goyal et al., 2015). Considering that only static models are used in the current filtering setup, it would be impossible to identify such effects in the docking results.

The remaining mutations are located outside of the ATP binding pocket. When looking at the data for resistance conferring mutations outside the pocket such as D761Y, L747P, L747S and L858R/V843I, it can be seen that the resistance conferred by these mutations is not correctly predicted. This is expected, since these mutations can only impact drug binding through their effects on the positioning of other residues that are located in the binding pocket. Since only static homology models were used, the extent to which such effects can be predicted is limited. This makes it unlikely that the effects of these mutations will be predicted correctly. Knowing this, it is important to realize that this approach is only effective when applied to mutations of those residues in close proximity to the drug binding pocket.



1) Activating mutations in the binding pocket.
2) Confers resistance by removing covalent interaction.
3) Confers resistance through increased protein flexibility.
4) Case handled at the UMCG tumorboard.
5) Mutation located outside of ATP binding pocket.

Figure 22 - Filtering results for EGFR mutations. First line for each mutant shows known inhibitor sensitivity (green: sensitive, red: resistant, gray: unknown), followed by a line of whitespace and the results per model. (4, 3, 2, 1 and 0 filter passes for green, yellow, orange, red and dark red respectively)

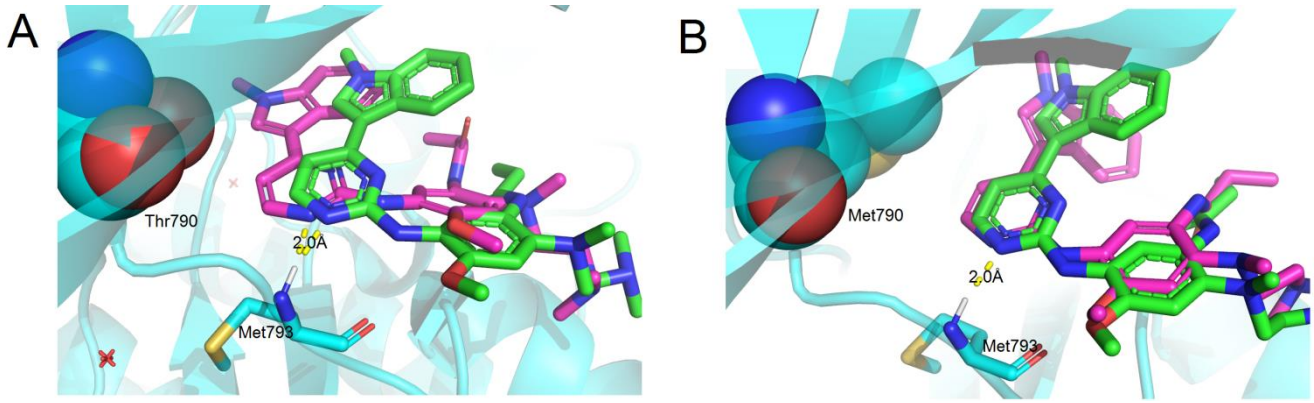


Figure 23 - A) Docking pose generated for osimertinib in the G719A model for the 1M17 template (shown in magenta) overlayed onto the crystal structure binding mode for osimertinib from the 4ZAU PDB structure (shown in green). B) Docking pose generated for osimertinib in the T790M model for the 1M17 template (shown in magenta) overlayed onto the crystal structure binding mode for osimertinib from the 4ZAU PDB structure (shown in green). The Thr790/Met790 residues are shown as space filling spheres. A small cleft of unoccupied space is located next to bound osimertinib in non-T790M models. During docking, this causes SMINA to shift the drug further into the binding pocket than is seen in the crystal structure binding mode. Nitrogens and oxygens are colored red and blue respectively

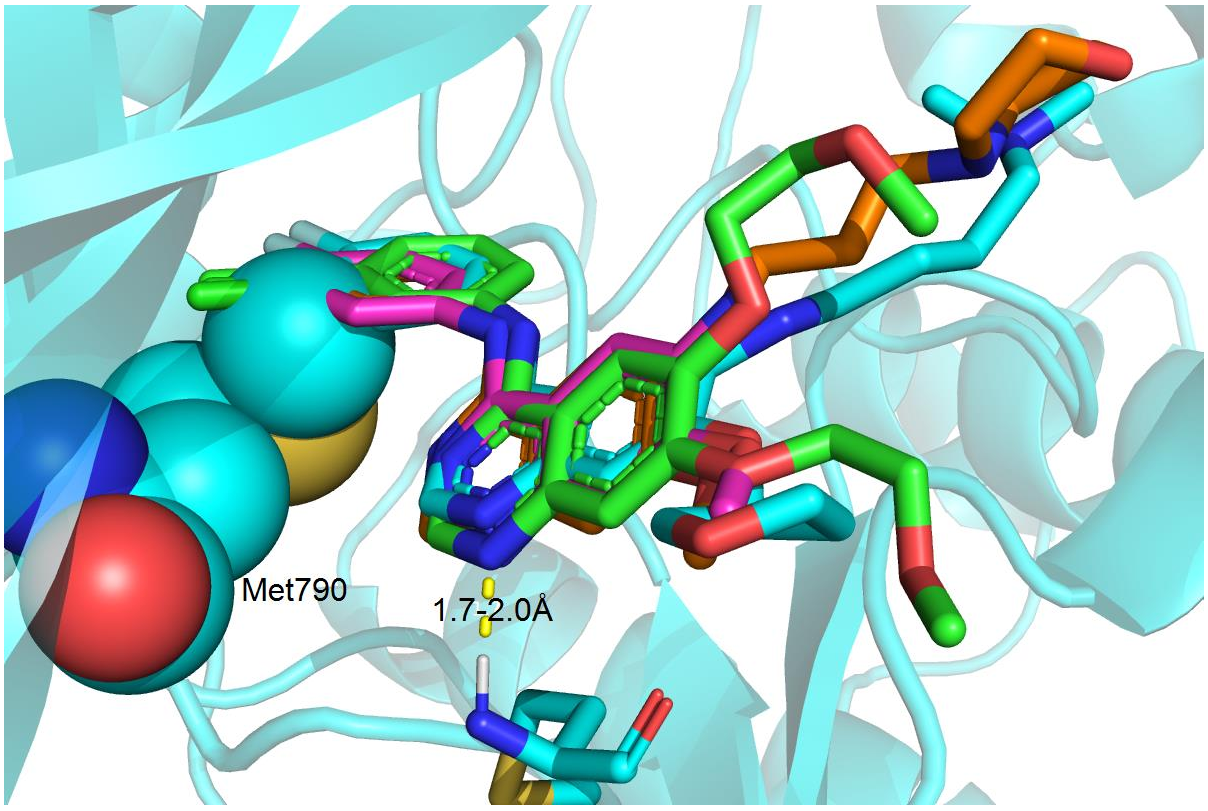


Figure 24 – Crystal structure binding modes for the EGFR inhibitors erlotinib, gefitinib, dacomitinib and afatinib overlayed onto the 4ZAU based homology model of T790M mutant EGFR. Clear steric interference with the known binding mode for these inhibitors can be seen in the image.

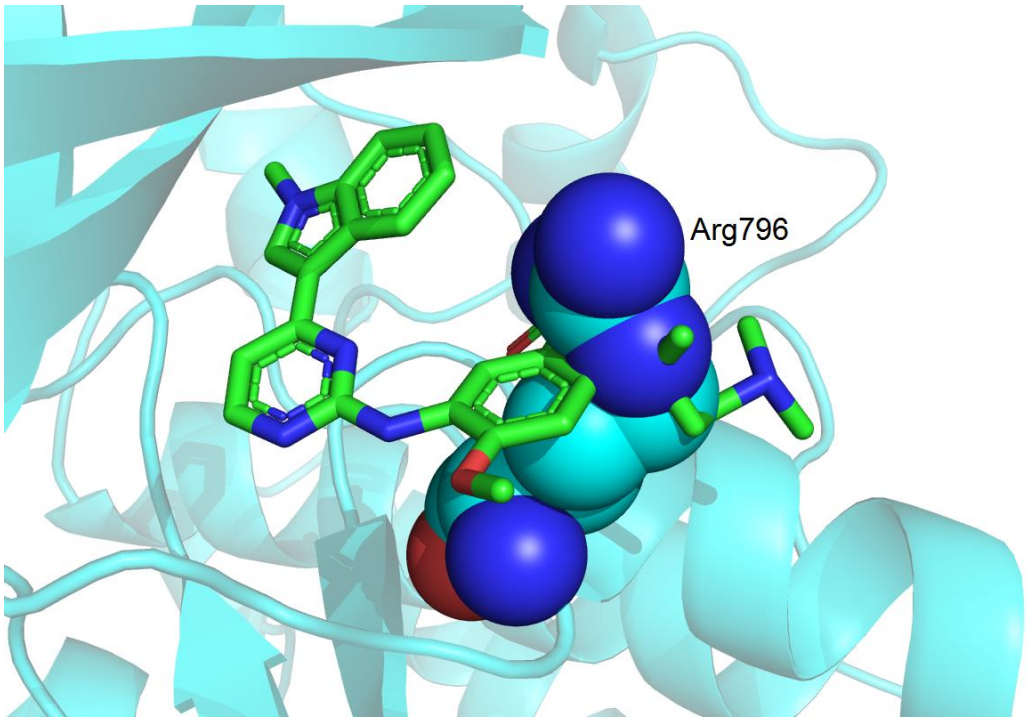


Figure 25 - Crystal structure binding modes for osimertinib overlayed onto the 4ZAU based homology model of T790M/G796R mutant EGFR. The mutated arginine residue in the SWISS-Model predicted orientation can be seen to take up a part of the pocket that osimertinib normally occupies when binding to EGFR.

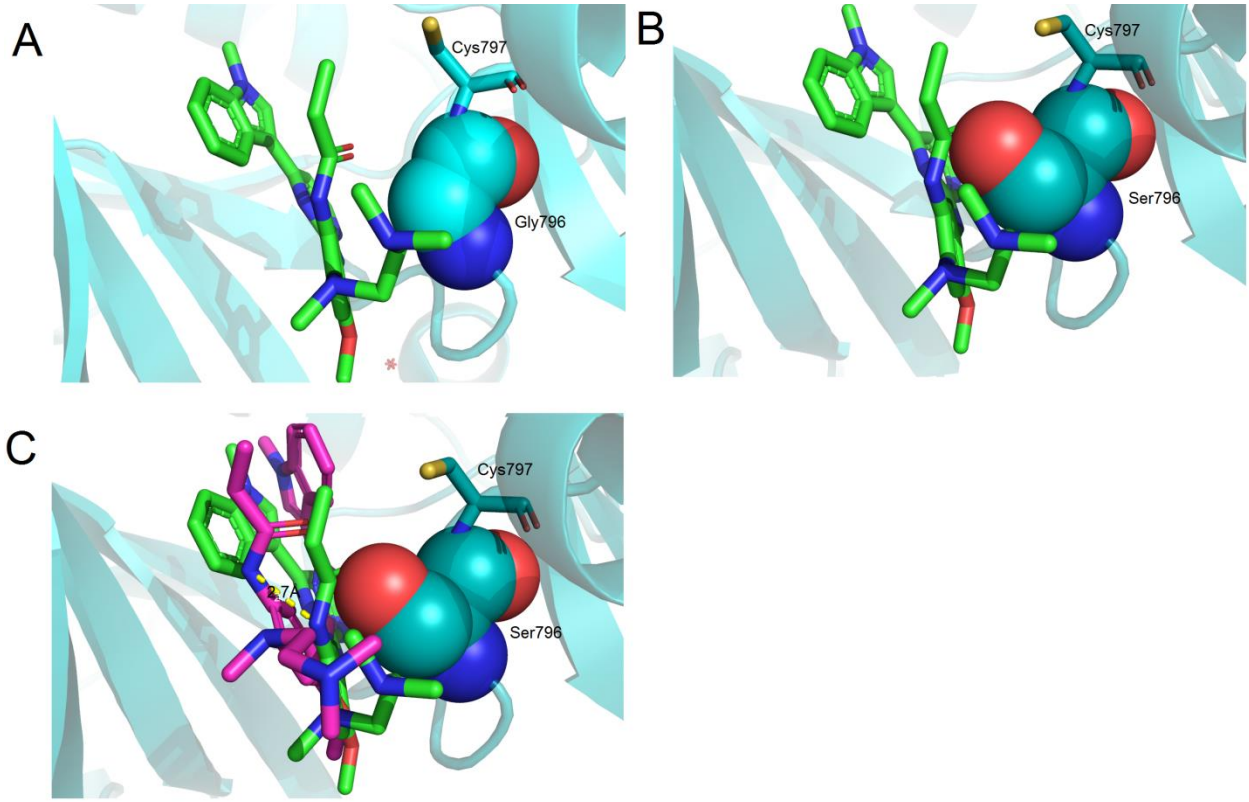


Figure 26 – A shows the crystal structure binding mode of osimertinib taken from the 4ZAU crystal structure. In B, it can be seen that the G796S mutation causes steric interference with the position of the acrylamide group that normally forms a covalent bond with the Cys797 residue. In C, A binding pose representative of those generated for osimertinib on G796S homology models is shown. Here, a displacement of the acrylamide linker of 1.7Å can be seen that is expected to prevent formation of the covalent bond.

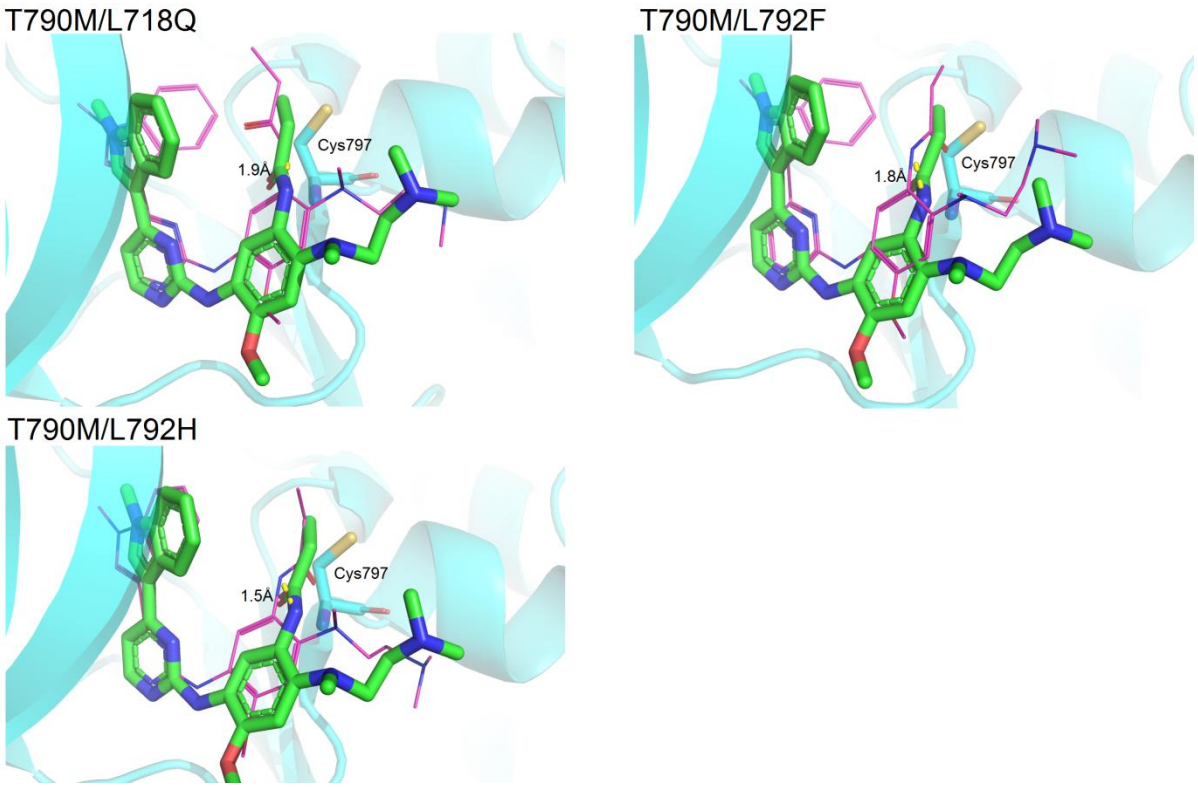


Figure 27 – Poses representative of those generated for various mutations that interfere with osimertinib binding. T790M/L718Q and T790M/L792F show a shift in the acrylamide linker group greater than the van de waals radius of a carbon (1.9Å and 1.8Å for both mutants respectively). This Is expected to hinder covalent bond formation. The T790M/L792H mutant showed a smaller shift in the acrylamide group of 1.5Å. All these mutations are known to confer resistance to osimertinib (Callegari et al., 2018; Y. Kobayashi et al., 2017; Q. Zhang et al., 2018)**. Nitrogens and oxygens are colored blue and red respectively.**

BRAF

For BRAF, the drugs dabrafenib and vemurafenib were docked on homology models based on PDB crystal structure 3OG7/4XV2 for V600 mutants and 4RZV/5CSW for other mutants. The filtering results for BRAF are shown in Figure 28.

The pdb structures 3og7 and 4xv2 were used in V600 mutant cases since they contain the V600E mutation, bringing them closer to target sequence of the homology model. The single V600 mutants are all either known or believed to be sensitive to BRAF inhibition (Chapman et al., 2011; Falchook et al., 2012). When looking at the docking results for these mutations, no convincing indicators of drug resistance can be seen. In V600K, one of the models fails to produce a pose passing all filters. When visually examining the generated poses however, it can be seen that a pose was generated which closely matches the crystal structure binding mode (Figure 29). While this pose comes just short of passing the hydrogen bond strength filter, its clear similarity with the crystal structure binding mode indicates that this does not indicate drug resistance. Because of this, the data for V600 mutations fits the inhibitor sensitivity described in literature.

While not having been reported in patients, it has been shown that mutations of the T529 residue, which are homologous to T790 mutations in EGFR, will likely cause drug resistance due to steric hindrance (Figure 30) (Whittaker et al., 2010). For this residue, homology models were created for the double mutants V600E/T529I, V600E/T529M and V600E/T529N. When looking at the filtering results for these mutations, it can be seen that in all cases, one or both models failed 3 or all filters. This indicates that the generated poses on these models strongly differ from the crystal structure

binding poses. Considering that the drugs docked properly on the crystal structures the models are based on, this indicates that these mutations could confer resistance against the BRAF inhibitors.

Another mutation in the binding pocket that is known to confer resistance to the BRAF inhibitor vemurafenib is V600K/L505H. In the docking results for this mutation, Both models failed to generate vemurafenib poses passing all filters. Visual inspection of the generated poses revealed that poses closely matching the crystal structure binding pose were generated in these models (Figure 31). However, while the poses themselves are similar, the SMINA predicted binding affinity shows a 17% increase in the binding affinity from -11.5kcal/mol to -9.5kcal/mol. This causes these poses to fail the binding affinity filter indicating that this mutation could confer resistance against vemurafenib.

A common mechanism of resistance against BRAF is increased dimerization with CRAF (Poulikakos et al., 2011). Since this a mechanism completely unrelated to inhibitor binding, it is expected that mutations with this mechanism will not show indicators of drug resistance in the filtering data. L514V/V600E is a mutation that likely confers resistance in this manner (J. Wang et al., 2018). As expected, when looking at the filtering results, no clear indication of drug resistance can be seen. The mutation G466V shares this mechanism of resistance as well (Wan et al., 2004). Here however, the data does indicate that the ability of the inhibitors to bind BRAF is affected. Visual inspection of the generated poses confirmed that none of the generated poses for dabrafenib on either model and no poses for vemurafenib on the 5CSW matched the crystal structure binding pose. Considering the mechanism of resistance however, the ability for the inhibitors to bind BRAF is not relevant in this case.

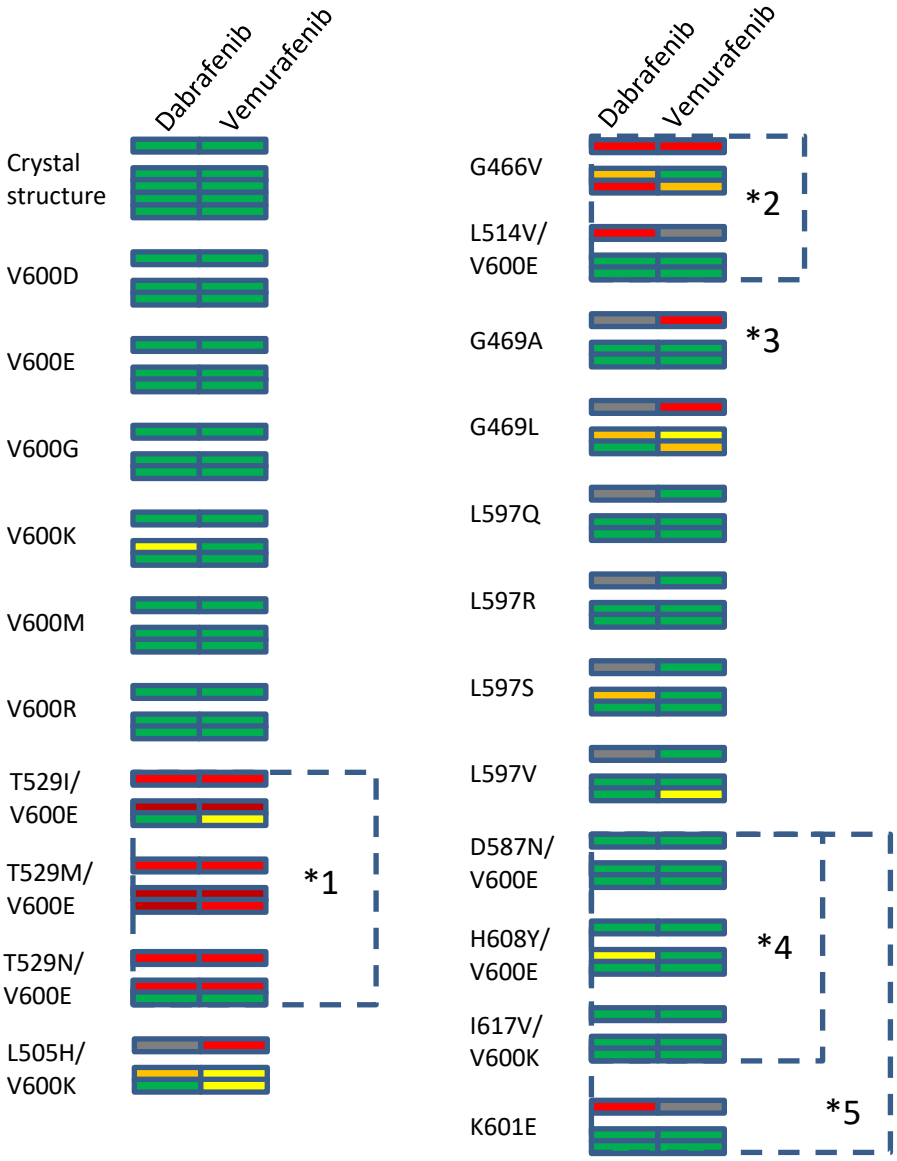
BRAF mutations of the G469 residue are known to lead to constitutively active BRAF (Porcelli et al., 2015). From case studies in which G469L mutant BRAF was expressed, it is known that this mutations is not sensitive to the inhibitor vemurafenib. This is believed to be the result of the leucine sidechain sterically interfering with the Lys483 residue. This causes the sidechain for the Lysine to shift closer to the binding pocket where it interferes with vemurafenib binding (Gautschi et al., 2013). When looking at the filtering results for this mutation, it can be seen that this data similarly suggests that this mutation confers resistance against vemurafenib. Poses for vemurafenib bound in both models either had a high RMSD or a predicted affinity that was greatly reduced relative to the docking results on the template. Like described by Gautschi et al, the homology models generated for docking also showed a small displacement of the Lys483 residue which could be responsible for the docking results obtained for this mutant (Figure 32).

The G469A has also been reported to cause resistance to vemurafenib (Porcelli et al., 2015). Contrary to G469L however, no indication of drug resistance can be seen in the filtering data for this mutant. A potential explanation for this result is that while G469A is resistance to a certain degree, vemurafenib can still bind and inhibit this mutant. This is known both from inhibition as well as patient data in which a partial response to vemurafenib was achieved in patients expressing this BRAF mutation.

The remaining mutations located in the binding pocket that have known sensitivity to BRAF inhibitors that were modeled are of the L597 residue. The BRAF mutants L597Q, L597R, L597S and L597V have all been reported to be sensitive to vemurafenib (Dahlman et al., 2012). When looking at the docking results for these mutations, it can be seen that they are in agreement with these results. Visual inspection of the passing poses revealed no clear differences with the crystal structure binding poses.

For BRAF, three mutations were modeled that were previously investigated at the UMCG tumorboard. These are V600E/D587N, V600E/H608Y and V600K/I617V. All of these were predicted to be sensitive to BRAF inhibition when analyzed manually. This was confirmed when the patients responded to BRAF inhibition therapy. In the filtering analysis, all models for V600E/D587N and V600K/I617V produced poses for dabrafenib and vemurafenib passing all filters. This indicates that these mutations are sensitive to BRAF inhibition which is in line with the patient outcomes. For V600E/H608Y, no pose passing all filters was found for dabrafenib in the 3OG7 model. Upon manual inspection of the lowest RMSD generated poses, a binding pose closely resembling the crystal structure pose for dabrafenib was found. This pose failed to pass all filters since its normalized h-bond strength was marginally lower than the cutoff value set. Considering how closely this pose matches the crystal structure binding mode, the results for the 3OG7 model can be said to predict that dabrafenib can properly bind this model. The pose in question is shown in Figure 33. Keeping that in mind, the results obtained using the filters match those found earlier in manual analysis.

Lastly, the out of pocket mutation K601E has been included. This mutation is known to be insensitive to dabrafenib treatment. In the filtering results however, no such effect is visible. Similar to EGFR, this shows the ineffectiveness of this setup when dealing with mutations not in close proximity to the bound drug.



1) Gatekeeper mutations equivalent to T790 mutantions in EGFR.
2) Confers resistance through promotion of BRAF/CRAF dimerization.
3) Case reports vary from no to partial response to vemurafenib.
4) Case handled at the UMCG tumorboard
5) Mutation located outside of ATP binding pocket

Figure 28 - Filtering results for BRAF mutations. First line for each mutant shows known inhibitor sensitivity (green: sensitive, red: resistant, gray: unknown), followed by a line of whitespace and the results per model. (4, 3, 2, 1 and 0 filter passes for green, yellow, orange, red and dark red respectively)

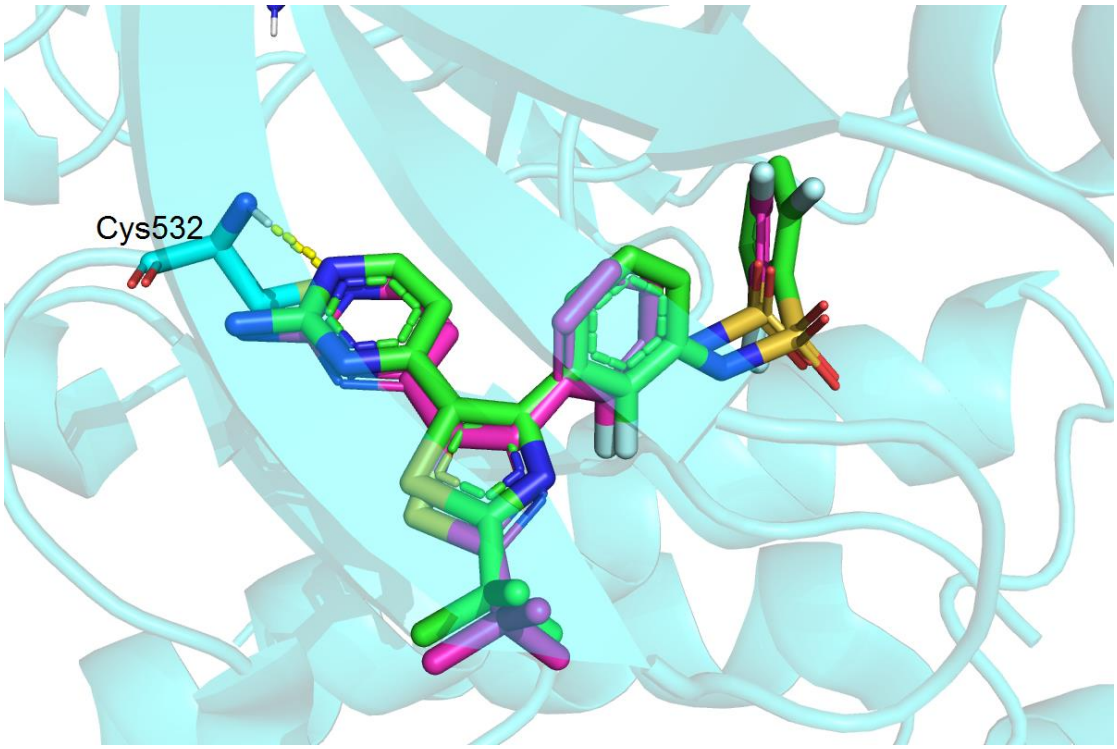


Figure 29 – Generated binding pose for dabrafenib on the 3OG7 model for V600K (shown in magenta). In the filtering setup, this pose fails to pass the hydrogen bond filter. Compared to the crystal structure binding pose for dabrafenib (shown in green), the generated pose appears highly similar, From this image it can be seen that this pose does accurately represent the crystal structure binding pose. This means that the failure to pass the hydrogen bond filter in this case does not indicate drug resistance.

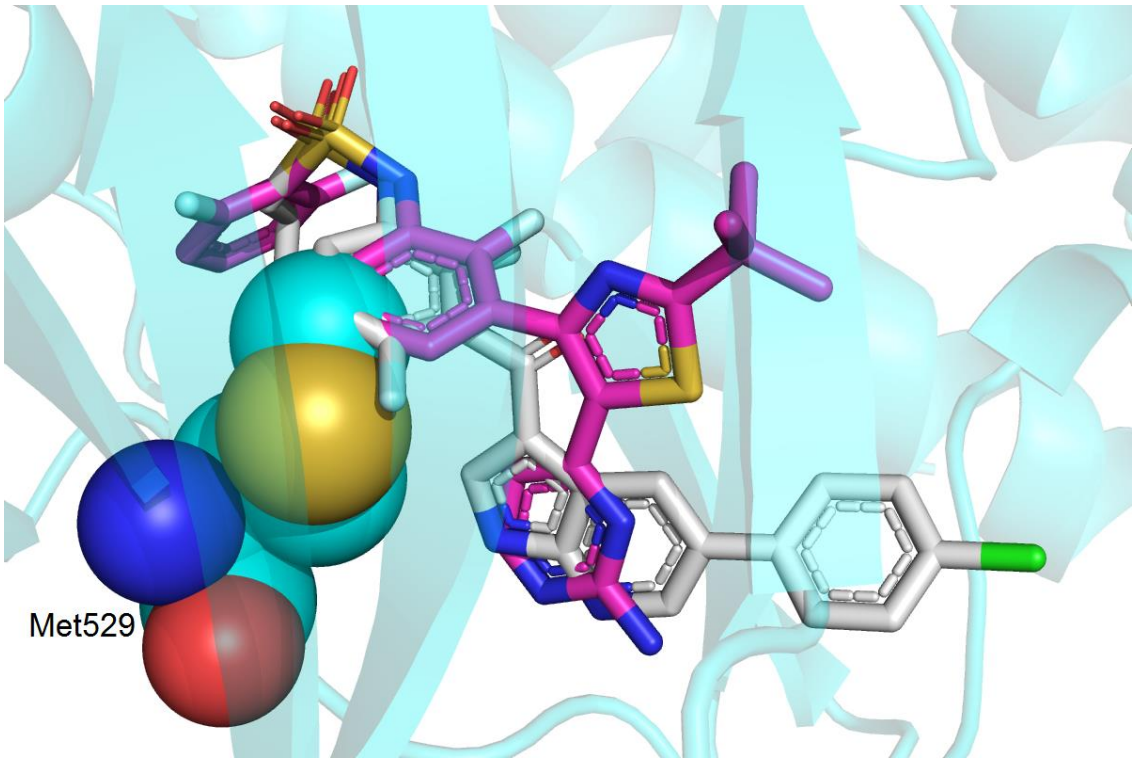


Figure 30 – Crystal structure binding poses of dabrafenib (magenta) and vemurafenib (white) superimposed on the V600E/T529M homology models for the 3OG7 template. As can be seen, the SWISS-Model predicted orientation of the methionine residue causes a steric clash with the crystal structure binding modes of the inhibitors.

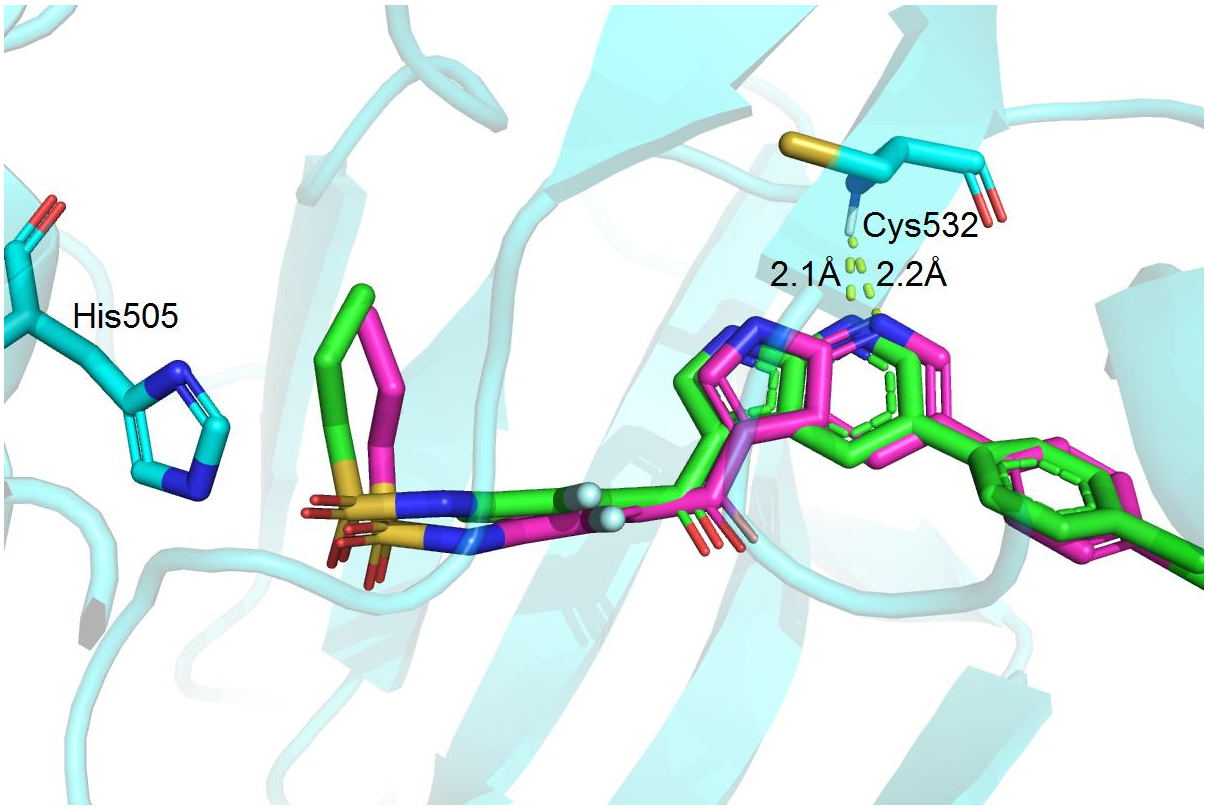


Figure 31- Vemurafenib pose generated on the 3OG7 model for V600K/L505H (shown in green). The pose can be seen to closely overlap with the crystal structure binding mode for vemurafenib found in the crystal structures (shown in magenta). However, the introduction of the polar Histidine residue caused a change in the predicted binding affinity from -11.5 kcal/mol to -9.5 kcal/mol. This is an increase of 17%, causing the pose to fail the affinity filter which can indicate drug resistance. Nitrogen, oxygen fluorine and sulfur are colored blue, red, white and orange respectively.

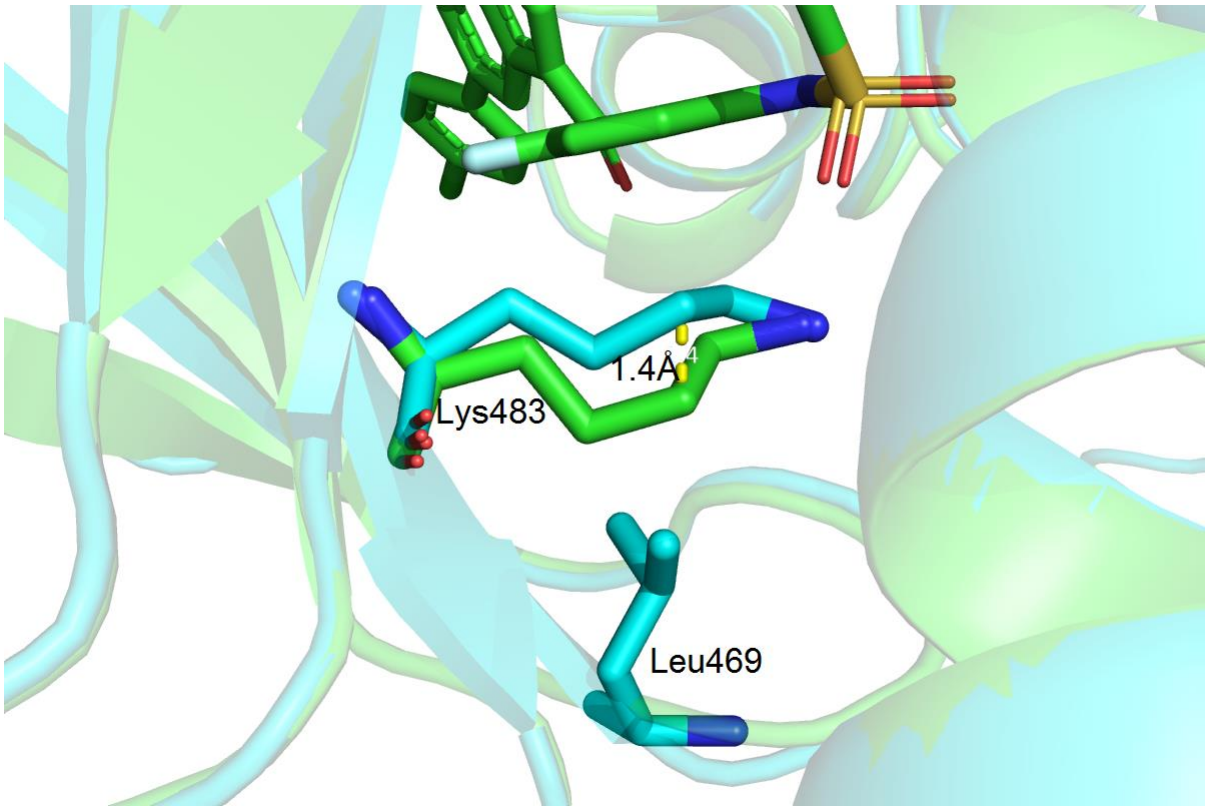


Figure 32 - Displacement of the BRAF L483sidechain resulting from the introduction of the G469L mutation in the 5CSW homology model for this mutation. This shift in the residue side chain is believed to prevent proper binding of vemurafenib in the binding pocket (Gautschi et al., 2013)**.**

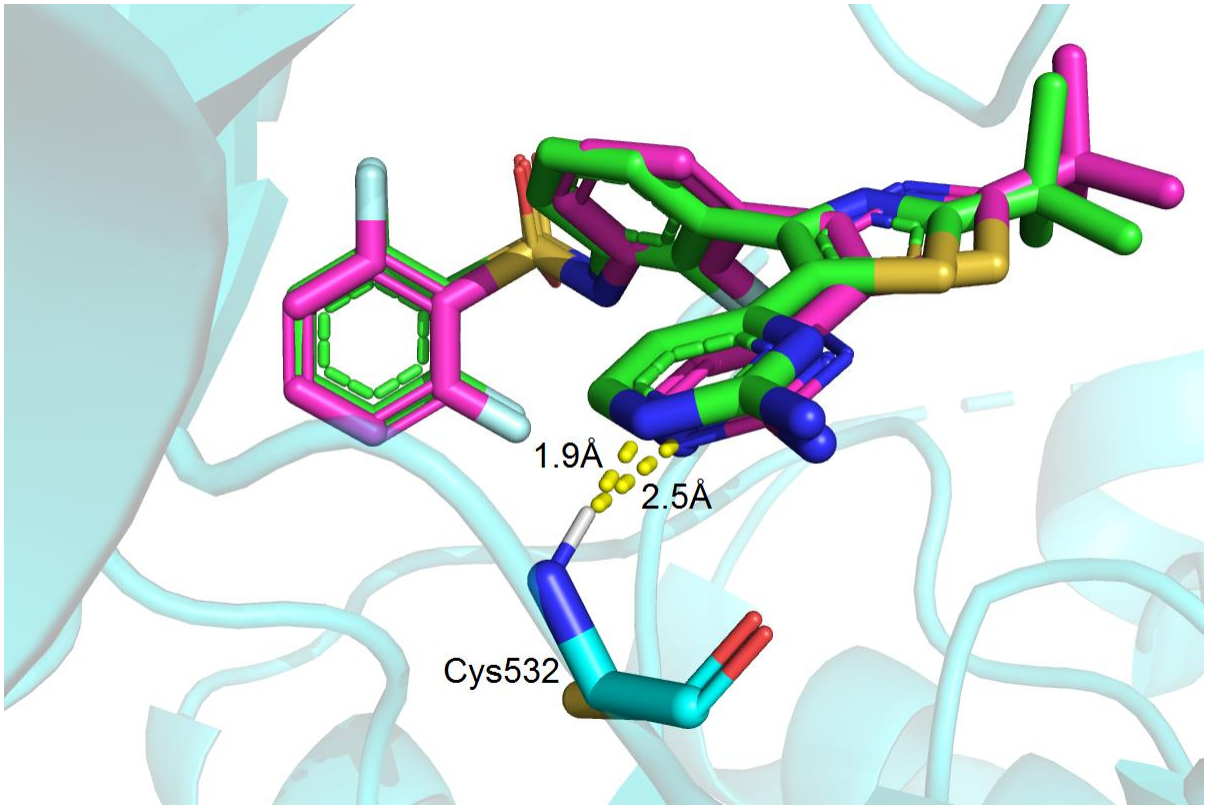


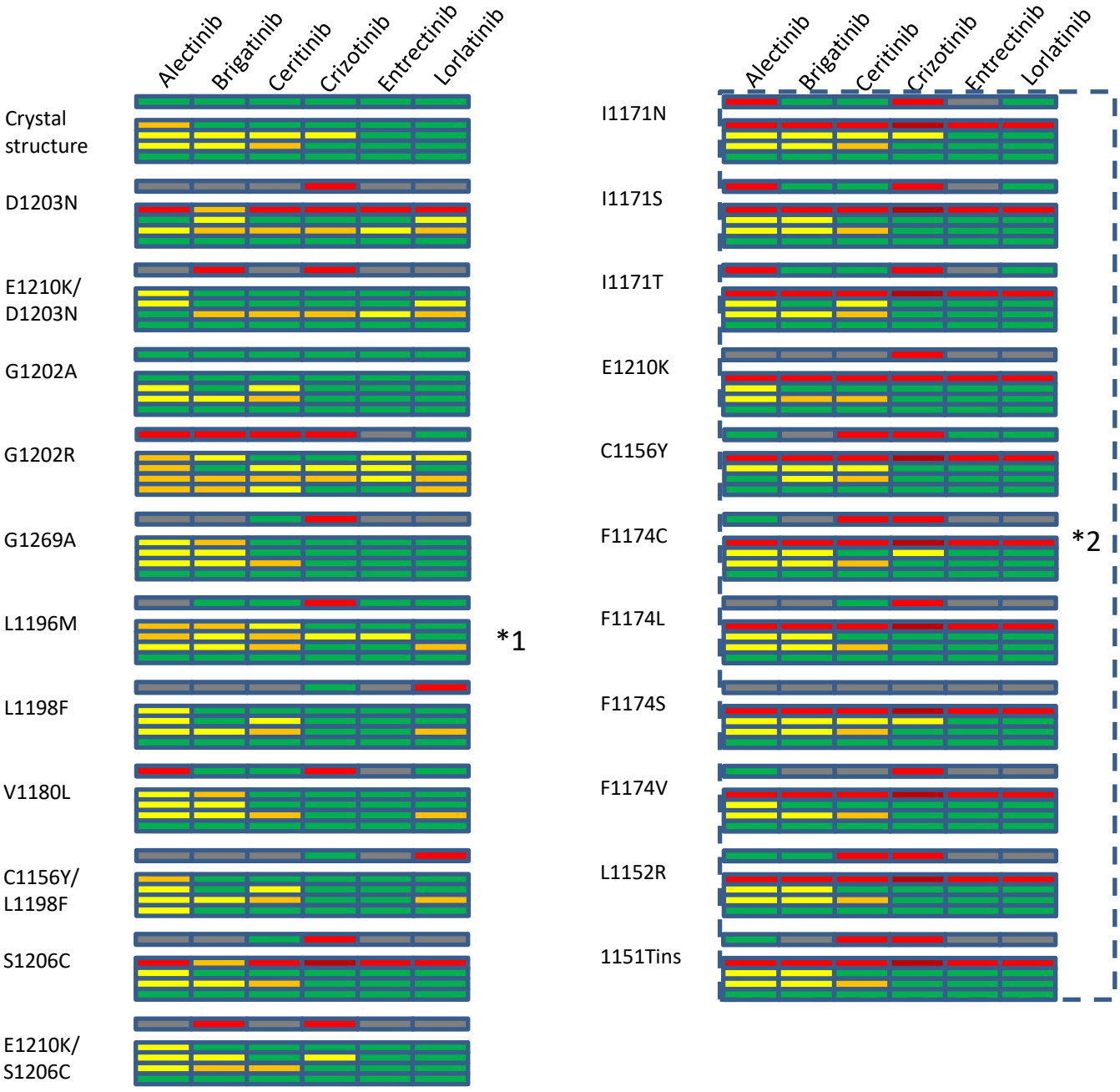
Figure 33 – Generated dabrafenib binding pose in the 3OG7 model for BRAF V600E/H608Y (shown in magenta). The crystal structure binding pose for dabrafenib is shown in green. It can be seen that the generated pose closely resembles the known binding mode of dabrafenib. However, his generated pose failed the hydrogen bond filter due to the length of

its hydrogen bond. Still, since the bond strength was only marginally lower than what was considered acceptable in the filter, this pose should be considered indicative of proper binding for dabrafenib in this model.

ALK

Similar to what was found during cross-docking, the filtering results indicate that many of the inhibitors did not dock properly on the template crystal structures. Since this indicates that the docking poses do not accurately reflect real-world binding events, the predictive capabilities for the filters using these models is expected to be very low.

Docking data was generated and filtered for the ALK mutations D1203N, E1210K/D1203N, G1202A, G1202R, G1269A, L1196M, L1198F, V1180L, C1156Y/L1198F, E1210K, E1210K/S1206C, I1171N, I1171S, I1171T, S1206C, C1156Y, F1174C, F1174L, F1174S, F1174V, L1152R and 1151Tins. The filtering results for ALK mutations are shown in Figure 34. In this figure, no clear correlation can be seen between the known drug resistance of the mutations and the number of filters passed. As an example. Figure 35 shows a generated docking pose for crizotinib docked onto the L1196M model for the 6MX8 template. The L1196M mutation is equivalent to the T790M mutation in EGFR, and should sterically interfere with the binding of crizotinib (Heuckmann et al., 2011). As can be seen in the image however, the generated crizotinib pose closely resembles the cristal structure binding mode. None of the docking results show any indication of crizotinib resistance caused by the L1196M mutation. Resistance conferred by other mutations aligns equally as poor with the filtering results. Based on this, it can be stated that the use of the current setup will provide little in terms of useful information for novel ALK mutations. A potential explanation for the poor performance of the filters in ALK cases is that this kinase is expressed in cancer solely due to a translocation of the gene. Many different translocations of ALK have been reported and these could have an effect on the structure of the ALK tyrosine kinase domain (Chiarle, Voena, Ambrogio, Piva, & Inghirami, 2008). The translocation of the gene therefore adds a variable affecting the structure of the kinase that is not taken into account in the docking setup. Since the effect of the fusion partner is not taken into account in the process, this could negatively impact the validity of the generated data. A possible reason for the poor performance of the filter in ALK cases is therefore the effect different fusion partners can have on the ALK structure.



1) Gatekeeper mutations equivalent to T790 mutantions in EGFR.
2) All mutations located outside of ATP binding pocket

Figure 34 - Filtering results for ALK mutations. First line for each mutant shows known inhibitor sensitivity (green: sensitive, red: resistant, gray: unknown), followed by a line of whitespace and the results per model. (4, 3, 2, 1 and 0 filter passes for green, yellow, orange, red and dark red respectively)

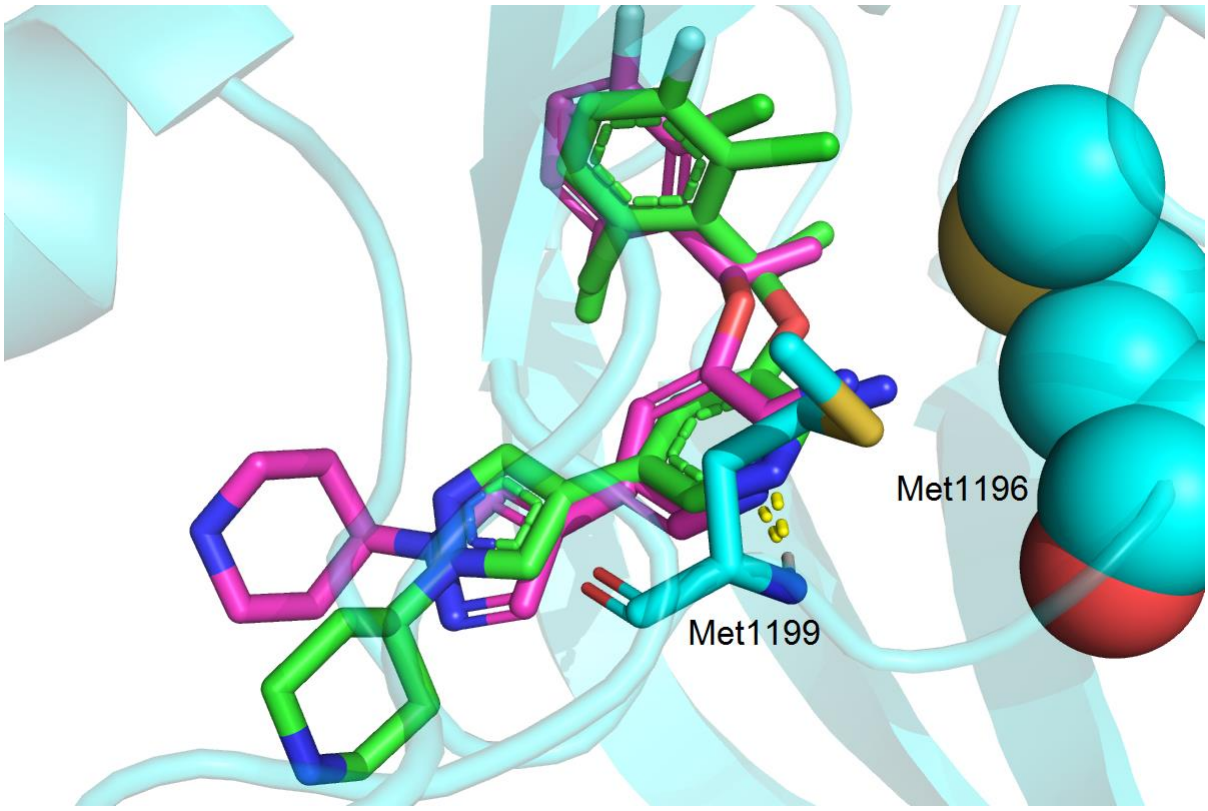


Figure 35 – Generated docking pose for cizotinib on the L1196M model for the 6MX8 template (shown in magenta). The generated pose can be seen to resemble the crystal structure binding pose for crizotinib shown in green. Nitrogen and oxygen are shown in blue and red respectively.

E. coli expression of ALK

In order to facilitate in future experiments for expressing the ALK tyrosine kinase domain in *E. coli,* an expression screening was performed. In the screening, *E. coli* cell lines BL21 Star (DE3), Origami 2(DE3), Rosetta 2(DE3)pLysS and LOBSTR were transformed with pETM-11 and pETM-30 vectors containing the DNA sequence for the human ALK tyrosine kinase domain in frame with the N-terminal his-tag. 1L of E. coli culture was produces as described in the materials and methods section after which the expression was assessed through SDS-PAGE on fractions collected during nickel-purification. The ALK tyrosine kinase domain has a size of roughly 37kDA. In expression with the pETM-11 vector, a band at this size could indicate that ALK is expressed. In the expression with pETM-30, the expression product is expected to have a size of around 64 kDa. The increase in size is due to the GST tag added by the vector. In the results for the pETM-11 vectors, a band corresponding to ALK expression was not seen in any of the cell-lines. While a line can be seen in the pellet that is close to the correct size, the same band can also be seen in the gels for the pETM-30 vector, indicating that this does not correspond to ALK.

For pETM-30 expression, the bands seen on the gel for the elution fractions were identical to those seen for pETM-11.This indicates that the lines seen do not correspond to expression of ALK. In the pellet fraction however, bands of the correct size were seen in the Origami 2 and LOBSTER cell lines. Based on these results, out of the strains tested, the Origami 2 and LOBSTER cell lines with the pETM-30 vector are seen as most suitable for ALK expression.

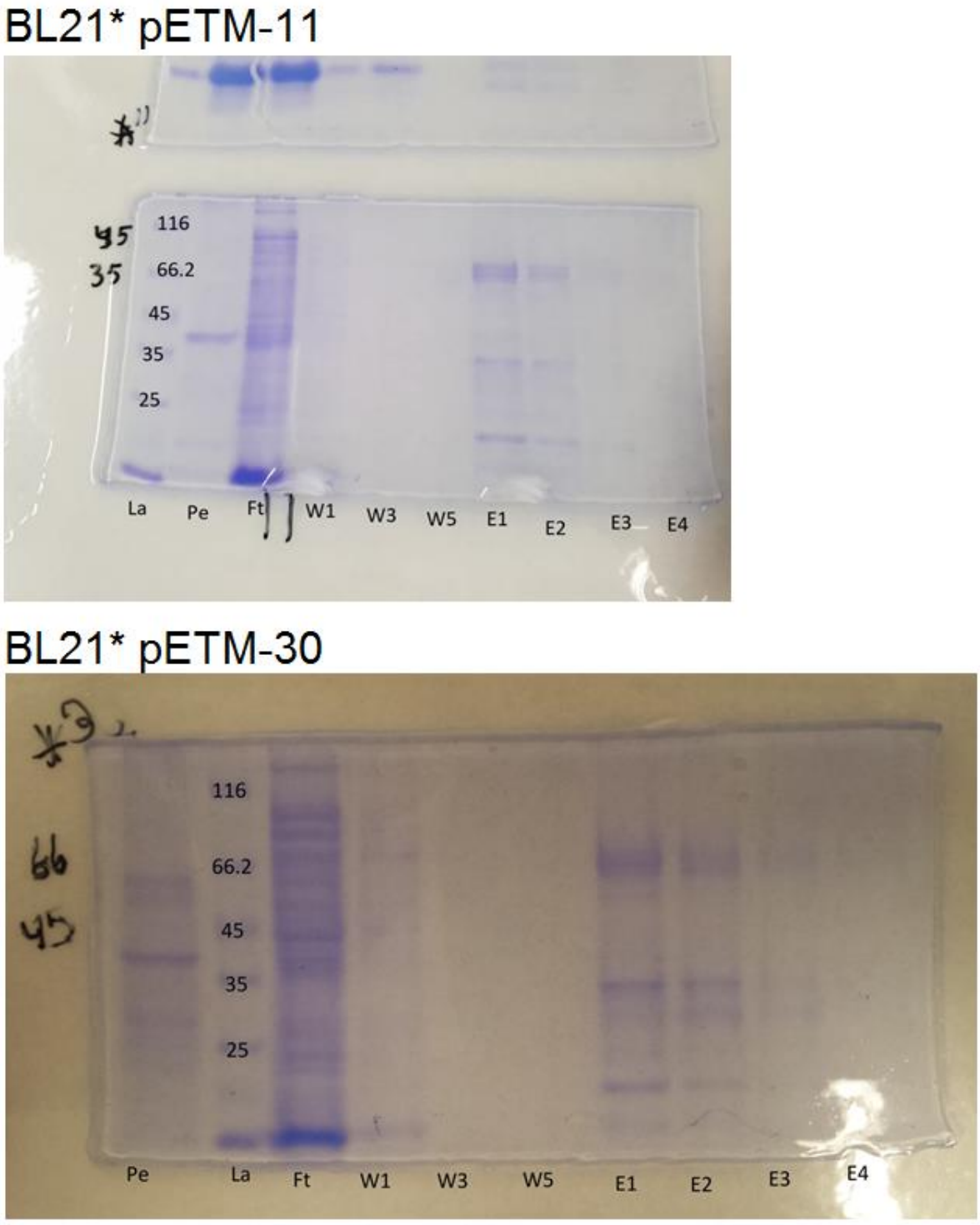


Figure 36 – SDS-PAGE gels for the BL21 Star (DE3) cell line transformed with the ALK-pETM-11 and ALK-pETM-30 construct. The lanes correspond to: Pe: Pellet (insoluble), La: Protein size ladder, Ft: Flowthrough, W…: Denotes a wash fraction, E…: Denotes an elution fraction.

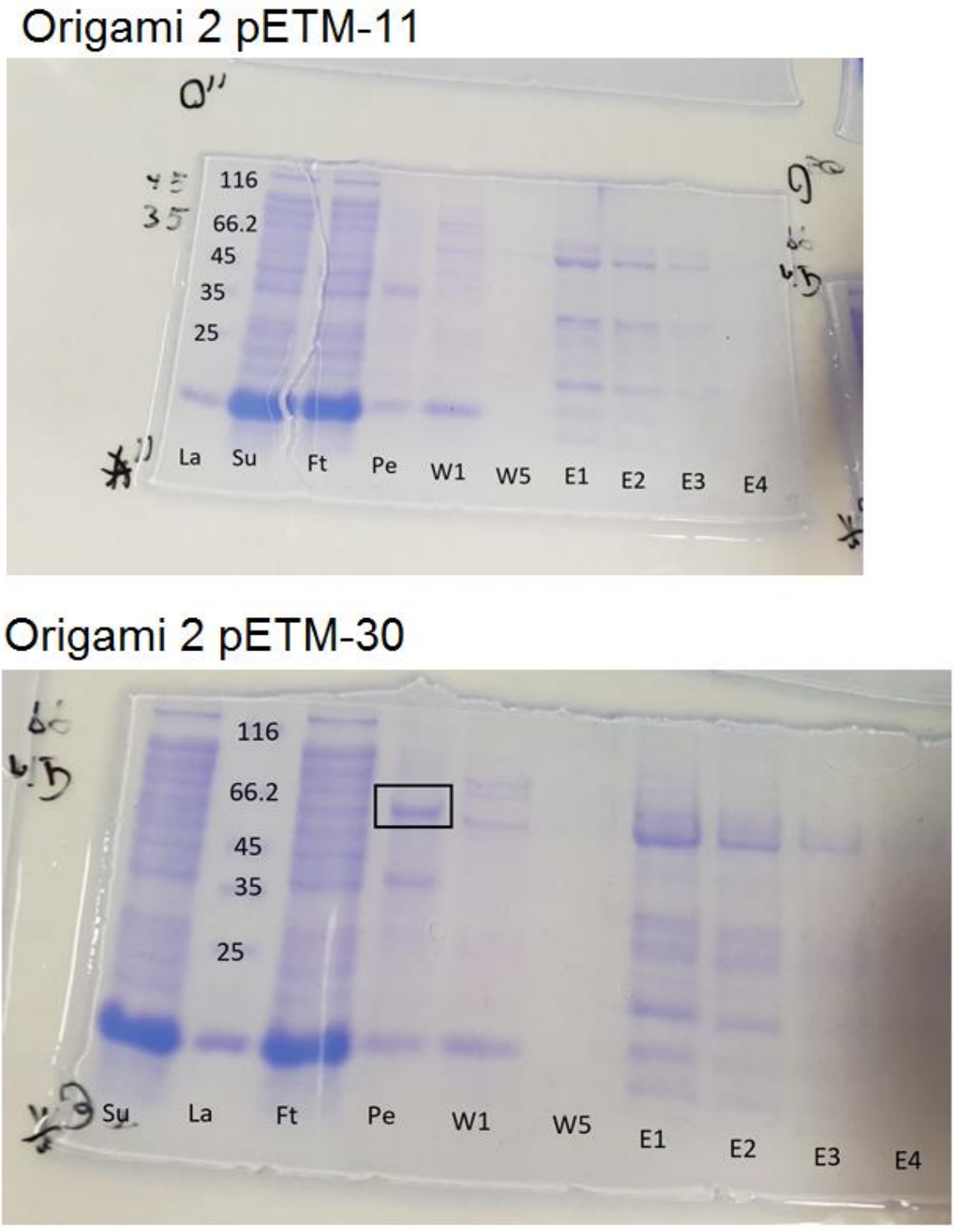


Figure 37 - SDS-PAGE gels for the Origami 2 (DE3) pLysS cell line transformed with the ALK-pETM-11 and ALK-pETM-30 construct. The lanes correspond to: Pe: Pellet (insoluble), La: Protein size ladder, Ft: Flowthrough, W…: Denotes a wash fraction, E…: Denotes an elution fraction. Expression of ALK was found in the pellet (insoluble)

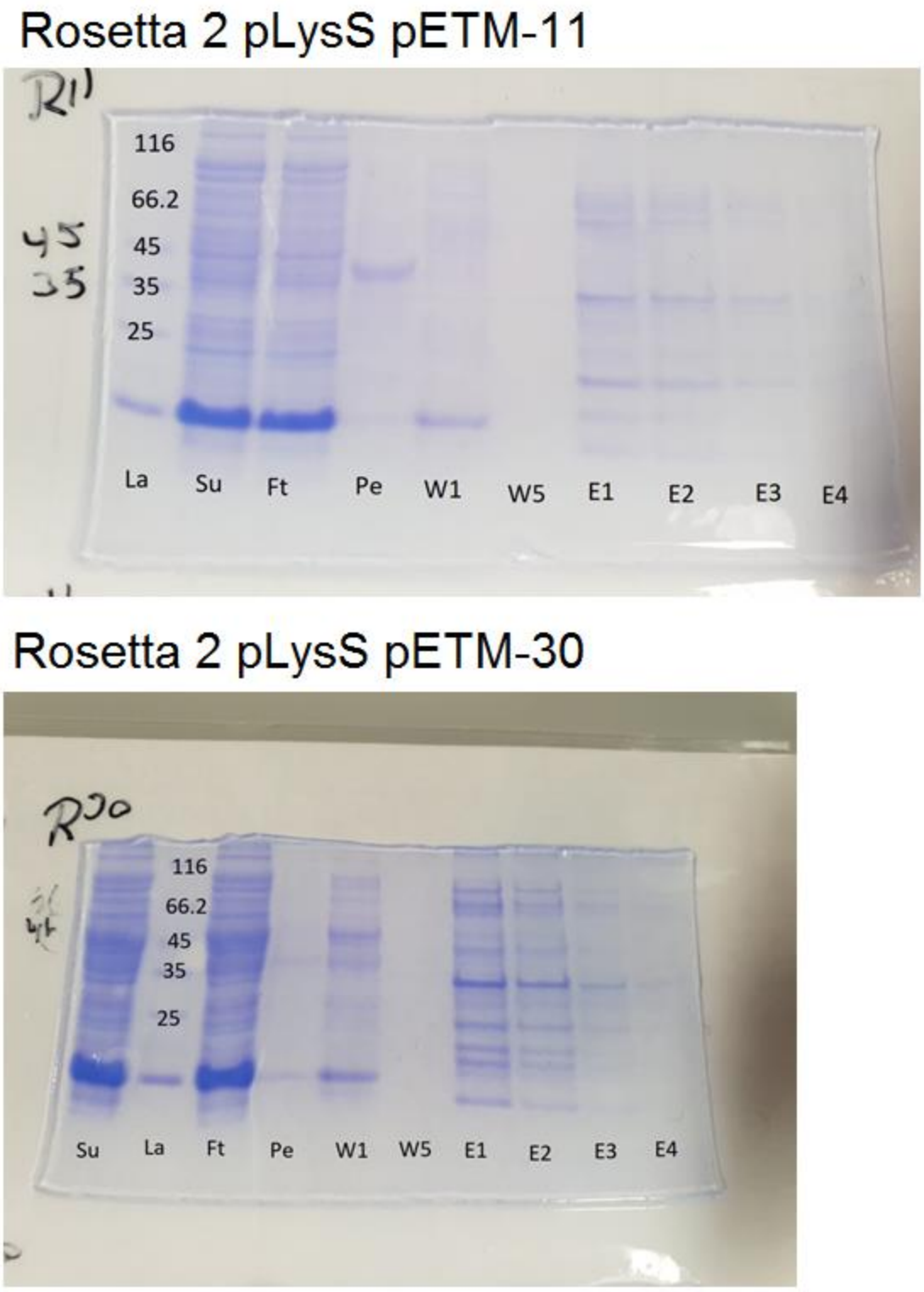


Figure 38 - SDS-PAGE gels for the Rosetta 2(DE3) cell line transformed with the ALK-pETM-11 and ALK-pETM-30 construct. The lanes correspond to: Pe: Pellet (insoluble), La: Protein size ladder, Ft: Flowthrough, W…: Denotes a wash fraction, E…: Denotes an elution fraction.

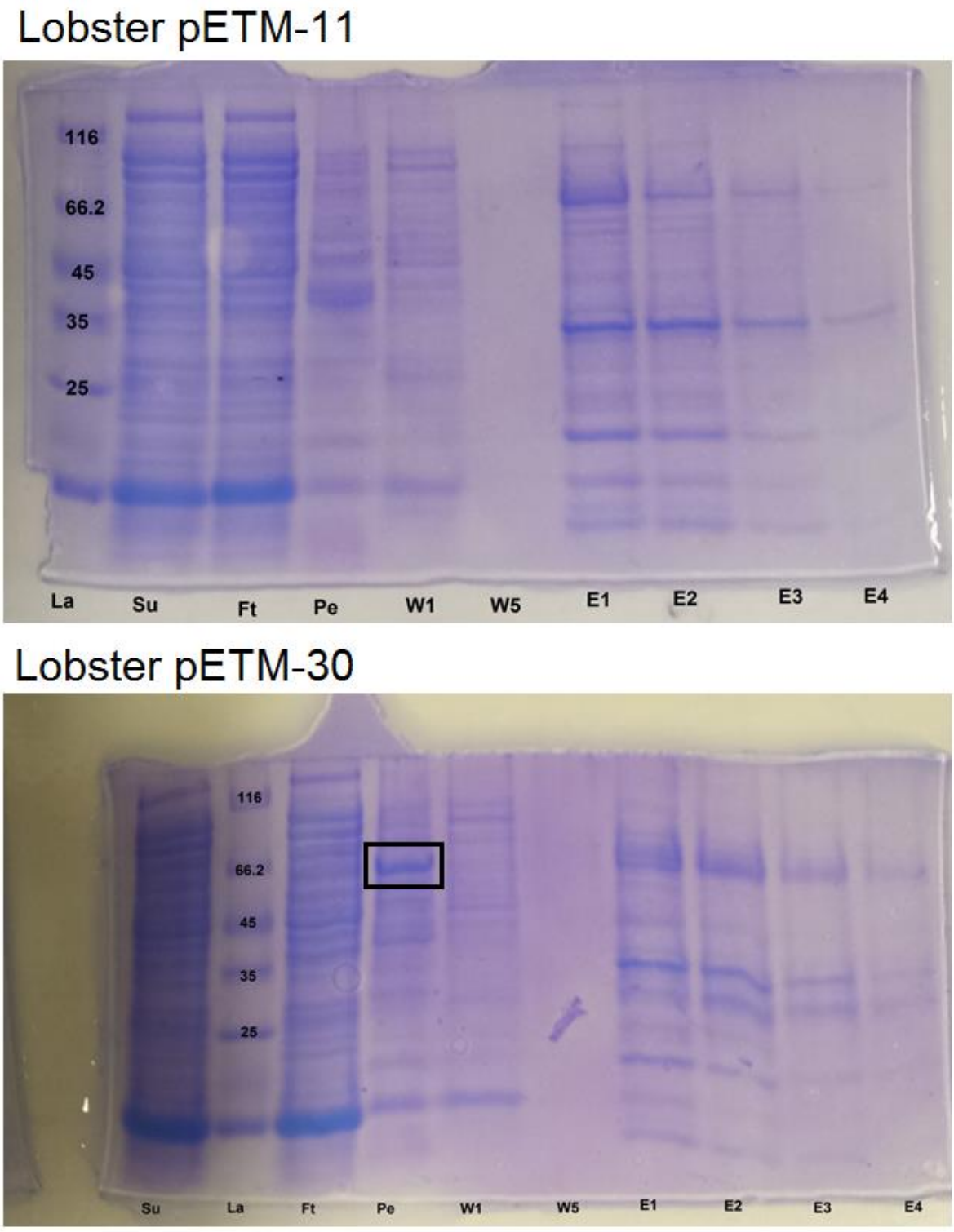


Figure 39 - SDS-PAGE gels for the LOBSTR cell line transformed with the ALK-pETM-11 and ALK-pETM-30 construct. The lanes correspond to: Pe: Pellet (insoluble), La: Protein size ladder, Ft: Flowthrough, W…: Denotes a wash fraction, E…: Denotes an elution fraction. Expression of ALK was found in the pellet (insoluble)

Conclusion

The use of molecular modeling and docking has much potential in predicting the efficacy of drugs. In kinase driven cancer, such computational methods have been shown to be able to generate useful information about the efficacy of kinase inhibitors in individual patients (van Kempen et al., 2018). Because of this, the implementation of such methods in clinical practice could enable medical professionals to make better informed decisions about patient treatment. Currently, limiting factors on the application of computational methods such as molecular docking are the time investment and knowledge needed to correctly interpret the generated data.

During this project, the use of docking filters to reduce the amount of docking data, as well as provide easily interpretable predictions on drug efficacy was explored in the context of kinase driven NSCLC. For EGFR and BRAF inhibitors, the use of such filters was shown to be able to identify generated binding poses matching the inhibitors binding mode as seen in crystal structures. This makes these filters a useful tool in reducing the volume of generated binding poses, allowing for easier interpretation of the data. Furthermore, when using such filters to score generated poses, it was shown that in many cases drug sensitivity could be correctly predicted in a fully automated fashion. In cases where the setup failed to provide clear predictions on drug efficacy, correct conclusions could in many cases still be drawn quickly by manually examining the lowest RMSD poses as identified using the filters. In cases where the conclusions based on the filter data were incorrect, the reasons for the filters to fail were usually explainable when considering the mutations mechanism of drug resistance.

Overall, this report has illustrated the use of molecular docking filters in predicting the efficacy of EGFR and BRAF inhibitors. This has much potential in allowing for easier and more widespread implementations of computational methods that can improve and personalized treatments in diseases such as NSCLC.

Akamine, T., Toyokawa, G., Tagawa, T., & Seto, T. (2018). Spotlight on lorlatinib and its potential in the treatment of NSCLC: the evidence to date. *OncoTargets and Therapy*, *Volume 11*, 5093–5101. https://doi.org/10.2147/OTT.S165511

Amaria, R. N., Prieto, P. A., Tetzlaff, M. T., Reuben, A., Andrews, M. C., Ross, M. I., … Wargo, J. A. (2018). Neoadjuvant plus adjuvant dabrafenib and trametinib versus standard of care in patients with high-risk, surgically resectable melanoma: a single-centre, open-label, randomised, phase 2 trial. *The Lancet Oncology*, *19*(2), 181–193. https://doi.org/10.1016/S1470-2045(18)30015-9

Amaro, R. E., Baudry, J., Chodera, J., Demir, Ö., McCammon, J. A., Miao, Y., & Smith, J. C. (2018). Ensemble Docking in Drug Discovery. *Biophysical Journal*, *114*(10), 2271–2278. https://doi.org/10.1016/j.bpj.2018.02.038

Balak, M. N., Gong, Y., Riely, G. J., Somwar, R., Li, A. R., Zakowski, M. F., … Pao, W. (2006). Novel D761Y and Common Secondary T790M Mutations in Epidermal Growth Factor Receptor-Mutant Lung Adenocarcinomas with Acquired Resistance to Kinase Inhibitors. *Clinical Cancer Research*, *12*(21), 6494–6501. https://doi.org/10.1158/1078-0432.CCR-06-1570

Barouch-Bentov, R., & Sauer, K. (2011). Mechanisms of drug resistance in kinases. *Expert Opinion on Investigational Drugs*, *20*(2), 153–208. https://doi.org/10.1517/13543784.2011.546344

Bayliss, R., Choi, J., Fennell, D. A., Fry, A. M., & Richards, M. W. (2016). Molecular mechanisms that underpin EML4-ALK driven cancers and their response to targeted drugs. *Cellular and Molecular Life Sciences*, *73*(6), 1209–1224. https://doi.org/10.1007/s00018-015-2117-6

Bean, J., Riely, G. J., Balak, M., Marks, J. L., Ladanyi, M., Miller, V. A., & Pao, W. (2008). Acquired Resistance to Epidermal Growth Factor Receptor Kinase Inhibitors Associated with a Novel T854A Mutation in a Patient with EGFR-Mutant Lung Adenocarcinoma. *Clinical Cancer Research*, *14*(22), 7519–7525. https://doi.org/10.1158/1078-0432.CCR-08-0151

Berman, H. M. (2000). The Protein Data Bank. *Nucleic Acids Research*, *28*(1), 235–242. https://doi.org/10.1093/nar/28.1.235

Bradish, J. R., & Cheng, L. (2014). Molecular pathology of malignant melanoma: changing the clinical practice paradigm toward a personalized approach. *Human Pathology*, *45*(7), 1315–1326. https://doi.org/10.1016/j.humpath.2014.04.001

Brose, M. S., Volpe, P., Feldman, M., Kumar, M., Rishi, I., Gerrero, R., … Weber, B. L. (2002). BRAF and RAS mutations in human lung cancer and melanoma. *Cancer Research*, *62*(23), 6997–7000. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12460918

Cakar, B., & Göker, E. (2016). Tyrosine Kinase Inhibitors. In *Breast Disease* (pp. 617–631). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-26012-9_36

Callegari, D., Ranaghan, K. E., Woods, C. J., Minari, R., Tiseo, M., Mor, M., … Lodola, A. (2018). L718Q mutant EGFR escapes covalent inhibition by stabilizing a non-reactive conformation of the lung cancer drug osimertinib. *Chemical Science*, *9*(10), 2740–2749. https://doi.org/10.1039/C7SC04761D

Carneiro, B. A., Pamarthy, S., Shah, A. N., Sagar, V., Unno, K., Han, H., … Abdulkadir, S. A. (2018). Anaplastic Lymphoma Kinase Mutation (ALK F1174C) in Small Cell Carcinoma of the Prostate and Molecular Response to Alectinib. *Clinical Cancer Research*, *24*(12), 2732–2739. https://doi.org/10.1158/1078-0432.CCR-18-0332

Chapman, P. B., Hauschild, A., Robert, C., Haanen, J. B., Ascierto, P., Larkin, J., … McArthur, G. A. (2011). Improved Survival with Vemurafenib in Melanoma with BRAF V600E Mutation. *New England Journal of Medicine*, *364*(26), 2507–2516. https://doi.org/10.1056/NEJMoa1103782

Chiarle, R., Voena, C., Ambrogio, C., Piva, R., & Inghirami, G. (2008). The anaplastic lymphoma kinase in the pathogenesis of cancer. *Nature Reviews Cancer*, *8*(1), 11–23. https://doi.org/10.1038/nrc2291

Chiba, M., Togashi, Y., Bannno, E., Kobayashi, Y., Nakamura, Y., Hayashi, H., … Nishio, K. (2017). Efficacy of irreversible EGFR-TKIs for the uncommon secondary resistant EGFR mutations L747S, D761Y, and T854A. *BMC Cancer*, *17*(1), 281. https://doi.org/10.1186/s12885-017-3263-z

Dahlman, K. B., Xia, J., Hutchinson, K., Ng, C., Hucks, D., Jia, P., … Pao, W. (2012). BRAF L597 Mutations in Melanoma Are Associated with Sensitivity to MEK Inhibitors. *Cancer Discovery*, *2*(9), 791–797. https://doi.org/10.1158/2159-8290.CD-12-0097

Dannenberg, J. J. (1998). An Introduction to Hydrogen Bonding By George A. Jeffrey (University of Pittsburgh). Oxford University Press: New York and Oxford. 1997. ix + 303 pp. $60.00. ISBN 0-19-509549-9. *Journal of the American Chemical Society*, *120*(22), 5604–5604. https://doi.org/10.1021/ja9756331

Davies, H., Bignell, G. R., Cox, C., Stephens, P., Edkins, S., Clegg, S., … Futreal, P. A. (2002). Mutations of the BRAF gene in human cancer. *Nature*, *417*(6892), 949–954. https://doi.org/10.1038/nature00766

Debruyne, D. N., Bhatnagar, N., Sharma, B., Luther, W., Moore, N. F., Cheung, N.-K., … George, R. E. (2016). ALK inhibitor resistance in ALKF1174L-driven neuroblastoma is associated with AXL activation and induction of EMT. *Oncogene*, *35*(28), 3681–3691. https://doi.org/10.1038/onc.2015.434

Drilon, A., Siena, S., Ou, S.-H. I., Patel, M., Ahn, M. J., Lee, J., … De Braud, F. G. (2017). Safety and Antitumor Activity of the Multitargeted Pan-TRK, ROS1, and ALK Inhibitor Entrectinib: Combined Results from Two Phase I Trials (ALKA-372-001 and STARTRK-1). *Cancer Discovery*, *7*(4), 400–409. https://doi.org/10.1158/2159-8290.CD-16-1237

Ercan, D., Choi, H. G., Yun, C.-H., Capelletti, M., Xie, T., Eck, M. J., … Janne, P. A. (2015). EGFR Mutations and Resistance to Irreversible Pyrimidine-Based EGFR Inhibitors. *Clinical Cancer Research*, *21*(17), 3913–3923. https://doi.org/10.1158/1078-0432.CCR-14-2789

Falchook, G. S., Long, G. V., Kurzrock, R., Kim, K. B., Arkenau, T. H., Brown, M. P., … Kefford, R. F. (2012). Dabrafenib in patients with melanoma, untreated brain metastases, and other solid tumours: a phase 1 dose-escalation trial. *The Lancet*, *379*(9829), 1893–1901. https://doi.org/10.1016/S0140-6736(12)60398-5

Fassunke, J., Müller, F., Keul, M., Michels, S., Dammert, M. A., Schmitt, A., … Sos, M. L. (2018). Overcoming EGFRG724S-mediated osimertinib resistance through unique binding characteristics of second-generation EGFR inhibitors. *Nature Communications*, *9*(1), 4655. https://doi.org/10.1038/s41467-018-07078-0

Ferguson, K. M. (2008). Structure-Based View of Epidermal Growth Factor Receptor Regulation. *Annual Review of Biophysics*, *37*(1), 353–373. https://doi.org/10.1146/annurev.biophys.37.032807.125829

Friboulet, L., Li, N., Katayama, R., Lee, C. C., Gainor, J. F., Crystal, A. S., … Engelman, J. A. (2014). The ALK Inhibitor Ceritinib Overcomes Crizotinib Resistance in Non-Small Cell Lung Cancer. *Cancer*

Discovery, *4*(6), 662–673. https://doi.org/10.1158/2159-8290.CD-13-0846

Fukuoka, M., Wu, Y.-L., Thongprasert, S., Sunpaweravong, P., Leong, S.-S., Sriuranpong, V., … Mok, T. S. K. (2011). Biomarker Analyses and Final Overall Survival Results From a Phase III, Randomized, Open-Label, First-Line Study of Gefitinib Versus Carboplatin/Paclitaxel in Clinically Selected Patients With Advanced Non–Small-Cell Lung Cancer in Asia (IPASS). *Journal of Clinical Oncology*, *29*(21), 2866–2874. https://doi.org/10.1200/JCO.2010.33.4235

Gainor, J. F., Dardaei, L., Yoda, S., Friboulet, L., Leshchiner, I., Katayama, R., … Shaw, A. T. (2016). Molecular Mechanisms of Resistance to First- and Second-Generation ALK Inhibitors in ALK-Rearranged Lung Cancer. *Cancer Discovery*, *6*(10), 1118–1133. https://doi.org/10.1158/2159-8290.CD-16-0596

Gautschi, O., Pauli, C., Strobel, K., Hirschmann, A., Printzen, G., Aebi, S., & Diebold, J. (2012). A Patient With BRAF V600E Lung Adenocarcinoma Responding to Vemurafenib. *Journal of Thoracic Oncology*, *7*(10), e23–e24. https://doi.org/10.1097/JTO.0b013e3182629903

Gautschi, O., Peters, S., Zoete, V., Aebersold-Keller, F., Strobel, K., Schwizer, B., … Diebold, J. (2013). Lung adenocarcinoma with BRAF G469L mutation refractory to vemurafenib. *Lung Cancer*, *82*(2), 365–367. https://doi.org/10.1016/j.lungcan.2013.08.012

Gazdar, A. F. (2009). Activating and resistance mutations of EGFR in non-small-cell lung cancer: role in clinical response to EGFR tyrosine kinase inhibitors. *Oncogene*, *28*(S1), S24–S31. https://doi.org/10.1038/onc.2009.198

Gerber, D. E., & Minna, J. D. (2010). ALK Inhibition for Non-Small Cell Lung Cancer: From Discovery to Therapy in Record Time. *Cancer Cell*, *18*(6), 548–551. https://doi.org/10.1016/j.ccr.2010.11.033

Goodsell, D. S. (2010). Epidermal Growth Factor Receptor. *RCSB Protein Data Bank*. https://doi.org/10.2210/rcsb_pdb/mom_2010_6

Goyal, S., Jamal, S., Shanker, A., & Grover, A. (2015). Structural investigations of T854A mutation in EGFR and identification of novel inhibitors using structure activity relationships. *BMC Genomics*, *16*(Suppl 5), S8. https://doi.org/10.1186/1471-2164-16-S5-S8

Grebner, C., Iegre, J., Ulander, J., Edman, K., Hogner, A., & Tyrchan, C. (2016). Binding Mode and Induced Fit Predictions for Prospective Computational Drug Design. *Journal of Chemical Information and Modeling*, *56*(4), 774–787. https://doi.org/10.1021/acs.jcim.5b00744

Gromiha, M. M., Nagarajan, R., & Selvaraj, S. (2019). Protein Structural Bioinformatics: An Overview. In *Encyclopedia of Bioinformatics and Computational Biology* (pp. 445–459). Elsevier. https://doi.org/10.1016/B978-0-12-809633-8.20278-1

Hammerschmidt, S., & Wirtz, H. (2009). Lung Cancer. *Deutsches Aerzteblatt Online*. https://doi.org/10.3238/arztebl.2009.0809

Han, S.-W., Kim, T.-Y., Hwang, P. G., Jeong, S., Kim, J., Choi, I. S., … Kim, N. K. (2005). Predictive and Prognostic Impact of Epidermal Growth Factor Receptor Mutation in Non–Small-Cell Lung Cancer Patients Treated With Gefitinib. *Journal of Clinical Oncology*, *23*(11), 2493–2501. https://doi.org/10.1200/JCO.2005.01.388

Hatcher, J. M., Bahcall, M., Choi, H. G., Gao, Y., Sim, T., George, R., … Gray, N. S. (2015). Discovery of Inhibitors That Overcome the G1202R Anaplastic Lymphoma Kinase Resistance Mutation. *Journal of Medicinal Chemistry*, *58*(23), 9296–9308. https://doi.org/10.1021/acs.jmedchem.5b01136

Heron, M. (2018). Deaths: Leading Causes for 2016. *National Vital Statistics Reports : From the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System*, *67*(6), 1–77. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/18092547

Heuckmann, J. M., Holzel, M., Sos, M. L., Heynck, S., Balke-Want, H., Koker, M., … Thomas, R. K. (2011). ALK Mutations Conferring Differential Resistance to Structurally Diverse ALK Inhibitors. *Clinical Cancer Research*, *17*(23), 7394–7401. https://doi.org/10.1158/1078-0432.CCR-11-1648

Holderfield, M., Deuker, M. M., McCormick, F., & McMahon, M. (2014). Targeting RAF kinases for cancer therapy: BRAF-mutated melanoma and beyond. *Nature Reviews Cancer*, *14*(7), 455–467. https://doi.org/10.1038/nrc3760

Hoogstraat, M., Gadellaa-van Hooijdonk, C. G., Ubink, I., Besselink, N. J. M., Pieterse, M., Veldhuis, W., … Lolkema, M. P. (2015). Detailed imaging and genetic analysis reveal a secondary BRAF L 505H resistance mutation and extensive intrapatient heterogeneity in metastatic BRAF mutant melanoma patients treated with vemurafenib. *Pigment Cell & Melanoma Research*, *28*(3), 318–323. https://doi.org/10.1111/pcmr.12347

Huang, S.-F. (2004). High Frequency of Epidermal Growth Factor Receptor Mutations with Complex Patterns in Non-Small Cell Lung Cancers Related to Gefitinib Responsiveness in Taiwan. *Clinical Cancer Research*, *10*(24), 8195–8203. https://doi.org/10.1158/1078-0432.CCR-04-1245

Irbäck, A., Mitternacht, S., & Mohanty, S. (2009). An effective all-atom potential for proteins. *BMC Biophysics*, *2*(1), 2. https://doi.org/10.1186/1757-5036-2-2

Jerabek-Willemsen, M., André, T., Wanner, R., Roth, H. M., Duhr, S., Baaske, P., & Breitsprecher, D. (2014). MicroScale Thermophoresis: Interaction analysis and beyond. *Journal of Molecular Structure*, *1077*, 101–113. https://doi.org/10.1016/j.molstruc.2014.03.009

Jones, J. E. (1924). On the Determination of Molecular Fields. II. From the Equation of State of a Gas. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *106*(738), 463–477. https://doi.org/10.1098/rspa.1924.0082

Katayama, R., Friboulet, L., Koike, S., Lockerman, E. L., Khan, T. M., Gainor, J. F., … Shaw, A. T. (2014). Two Novel ALK Mutations Mediate Acquired Resistance to the Next-Generation ALK Inhibitor Alectinib. *Clinical Cancer Research*, *20*(22), 5686–5696. https://doi.org/10.1158/1078-0432.CCR-14-1511

Katayama, Ryohei. (2018, March). Drug resistance in anaplastic lymphoma kinase-rearranged lung cancer. *Cancer Science*. https://doi.org/10.1111/cas.13504

Kobayashi, S., Boggon, T. J., Dayaram, T., Jänne, P. A., Kocher, O., Meyerson, M., … Halmos, B. (2005). EGFR Mutation and Resistance of Non–Small-Cell Lung Cancer to Gefitinib. *New England Journal of Medicine*, *352*(8), 786–792. https://doi.org/10.1056/NEJMoa044238

Kobayashi, Y., Azuma, K., Nagai, H., Kim, Y. H., Togashi, Y., Sesumi, Y., … Mitsudomi, T. (2017). Characterization of EGFR T790M, L792F, and C797S Mutations as Mechanisms of Acquired Resistance to Afatinib in Lung Cancer. *Molecular Cancer Therapeutics*, *16*(2), 357–364. https://doi.org/10.1158/1535-7163.MCT-16-0407

Kobayashi, Y., Togashi, Y., Yatabe, Y., Mizuuchi, H., Jangchul, P., Kondo, C., … Mitsudomi, T. (2015). EGFR Exon 18 Mutations in Lung Cancer: Molecular Predictors of Augmented Sensitivity to Afatinib or Neratinib as Compared with First- or Third-Generation TKIs. *Clinical Cancer Research*, *21*(23), 5305–5313. https://doi.org/10.1158/1078-0432.CCR-15-1046

Koes, D. R., Baumgartner, M. P., & Camacho, C. J. (2013). Lessons Learned in Empirical Scoring with smina from the CSAR 2011 Benchmarking Exercise. *Journal of Chemical Information and Modeling*, *53*(8), 1893–1904. https://doi.org/10.1021/ci300604z

Kris, M. G., Johnson, B. E., Kwiatkowski, D. J., Iafrate, A. J., Wistuba, I. I., Aronson, S. L., … Bunn, P. A. (2011). Identification of driver mutations in tumor specimens from 1,000 patients with lung adenocarcinoma: The NCI's Lung Cancer Mutation Consortium (LCMC). *Journal of Clinical Oncology*, *29*(15_suppl), CRA7506–CRA7506. https://doi.org/10.1200/jco.2011.29.15_suppl.cra7506

Kumar, A., Petri, E. T., Halmos, B., & Boggon, T. J. (2008). Structure and Clinical Relevance of the Epidermal Growth Factor Receptor in Human Cancer. *Journal of Clinical Oncology*, *26*(10), 1742–1751. https://doi.org/10.1200/JCO.2007.12.1178

Kumar, R., Schmidt, J. R., & Skinner, J. L. (2007). Hydrogen bonding definitions and dynamics in liquid water. *The Journal of Chemical Physics*, *126*(20), 204107. https://doi.org/10.1063/1.2742385

Larkin, J., & Fisher. (2012). Vemurafenib: a new treatment for BRAF-V600 mutated advanced melanoma. *Cancer Management and Research*, 243. https://doi.org/10.2147/CMAR.S25284

Lemmon, M. A., & Schlessinger, J. (2010). Cell Signaling by Receptor Tyrosine Kinases. *Cell*, *141*(7), 1117–1134. https://doi.org/10.1016/j.cell.2010.06.011

Li, J., Sun, R., Wu, Y., Song, M., Li, J., Yang, Q., … Zhao, Q. (2017). L1198F Mutation Resensitizes Crizotinib to ALK by Altering the Conformation of Inhibitor and ATP Binding Sites. *International Journal of Molecular Sciences*, *18*(3), 482. https://doi.org/10.3390/ijms18030482

Lin, J. J., Zhu, V. W., Schoenfeld, A. J., Yeap, B. Y., Saxena, A., Ferris, L. A., … Riely, G. J. (2018). Brigatinib in Patients With Alectinib-Refractory ALK-Positive NSCLC. *Journal of Thoracic Oncology*, *13*(10), 1530–1538. https://doi.org/10.1016/j.jtho.2018.06.005

Luk, P. P., Yu, B., Ng, C. C., Mercorella, B., Selinger, C., Lum, T., … Cooper, W. A. (2015). BRAF mutations in non-small cell lung cancer. *Translational Lung Cancer Research*, *4*(2), 142–148. https://doi.org/10.3978/j.issn.2218-6751.2014.08.08

Lynch, T. J., Bell, D. W., Sordella, R., Gurubhagavatula, S., Okimoto, R. A., Brannigan, B. W., … Haber, D. A. (2004). Activating Mutations in the Epidermal Growth Factor Receptor Underlying Responsiveness of Non–Small-Cell Lung Cancer to Gefitinib. *New England Journal of Medicine*, *350*(21), 2129–2139. https://doi.org/10.1056/NEJMoa040938

Morando, M. A., Saladino, G., D'Amelio, N., Pucheta-Martinez, E., Lovera, S., Lelli, M., … Gervasio, F. L. (2016). Conformational Selection and Induced Fit Mechanisms in the Binding of an Anticancer Drug to the c-Src Kinase. *Scientific Reports*, *6*(1), 24439. https://doi.org/10.1038/srep24439

Murali, R., Menzies, & Long, G. (2012). Dabrafenib and its potential for the treatment of metastatic melanoma. *Drug Design, Development and Therapy*, *6*, 391. https://doi.org/10.2147/DDDT.S38998

Nagar, B., Bornmann, W. G., Pellicena, P., Schindler, T., Veach, D. R., Miller, W. T., … Kuriyan, J. (2002). Crystal structures of the kinase domain of c-Abl in complex with the small molecule inhibitors PD173955 and imatinib (STI-571). *Cancer Research*, *62*(15), 4236–4243. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12154025

Nasu, S., Shiroyama, T., Morita, S., Takata, S., Takada, H., Masuhiro, K., … Hirashima, T. (2018). Osimertinib Treatment Was Unsuccessful for Lung Adenocarcinoma with G719S, S768I, and

T790M Mutations. *Internal Medicine*, *57*(24), 3643–3645. https://doi.org/10.2169/internalmedicine.0923-18

Normanno, N., De Luca, A., Bianco, C., Strizzi, L., Mancino, M., Maiello, M. R., … Salomon, D. S. (2006). Epidermal growth factor receptor (EGFR) signaling in cancer. *Gene*, *366*(1), 2–16. https://doi.org/10.1016/j.gene.2005.10.018

Oda, K., Matsuoka, Y., Funahashi, A., & Kitano, H. (2005). A comprehensive pathway map of epidermal growth factor receptor signaling. *Molecular Systems Biology*, *1*, 2005.0010. https://doi.org/10.1038/msb4100014

Ou, S.-H. I., Cui, J., Schrock, A. B., Goldberg, M. E., Zhu, V. W., Albacker, L., … Ali, S. M. (2017). Emergence of novel and dominant acquired EGFR solvent-front mutations at Gly796 (G796S/R) together with C797S/R and L792F/H mutations in one EGFR (L858R/T790M) NSCLC patient who progressed on osimertinib. *Lung Cancer*, *108*(April), 228–231. https://doi.org/10.1016/j.lungcan.2017.04.003

Ou, S. H., Milliken, J. C., Azada, M. C., Miller, V. A., Ali, S. M., & Klempner, S. J. (2016). ALK F1174V mutation confers sensitivity while ALK I1171 mutation confers resistance to alectinib. The importance of serial biopsy post progression. *Lung Cancer*, *91*(2016), 70–72. https://doi.org/10.1016/j.lungcan.2015.09.006

Oztan, A., Fischer, S., Schrock, A. B., Erlich, R. L., Lovly, C. M., Stephens, P. J., … Raez, L. E. (2017). Emergence of EGFR G724S mutation in EGFR-mutant lung adenocarcinoma post progression on osimertinib. *Lung Cancer*, *111*(July), 84–87. https://doi.org/10.1016/j.lungcan.2017.07.002

Paul, M. K., & Mukhopadhyay, A. K. (2004). Tyrosine kinase - Role and significance in Cancer. *International Journal of Medical Sciences*, *1*(2), 101–115. https://doi.org/10.7150/ijms.1.101

Peyssonnaux, C., & Eychène, A. (2001). The Raf/MEK/ERK pathway: new concepts of activation. *Biology of the Cell*, *93*(1–2), 53–62. https://doi.org/10.1016/S0248-4900(01)01125-X

Porcelli, L., Guida, G., Cocco, T., Quatrale, A. E., Iacobazzi, R. M., Stolfa, D. A., … Azzariti, A. (2015). The relevance of BRAF G469A mutation in determining the response to therapy in metastatic melanoma. *Journal of Translational Medicine*, *13*(Suppl 1), P3. https://doi.org/10.1186/1479-5876-13-s1-p3

Poulikakos, P. I., Persaud, Y., Janakiraman, M., Kong, X., Ng, C., Moriceau, G., … Solit, D. B. (2011). RAF inhibitor resistance is mediated by dimerization of aberrantly spliced BRAF(V600E). *Nature*, *480*(7377), 387–390. https://doi.org/10.1038/nature10662

Ravindranath, P. A., Forli, S., Goodsell, D. S., Olson, A. J., & Sanner, M. F. (2015). AutoDockFR: Advances in Protein-Ligand Docking with Explicitly Specified Binding Site Flexibility. *PLOS Computational Biology*, *11*(12), e1004586. https://doi.org/10.1371/journal.pcbi.1004586

Rolfo, C., Passiglia, F., Castiglia, M., Raez, L. E., Germonpre, P., Zwaenepoel, K., … Pauwels, P. (2014). ALK and crizotinib : after the honeymoon … what else ? Resistance mechanisms and new therapies to overcome it. *Lung Cancer Research*, *3*(4), 250–261. https://doi.org/10.3978/j.issn.2218-6751.2014.03.01

Rosell, R., Carcereny, E., Gervais, R., Vergnenegre, A., Massuti, B., Felip, E., … Paz-Ares, L. (2012). Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial. *The Lancet Oncology*, *13*(3), 239–246. https://doi.org/10.1016/S1470-2045(11)70393-X

Roskoski, R. (2013). Anaplastic lymphoma kinase (ALK): Structure, oncogenic activation, and pharmacological inhibition. *Pharmacological Research*. https://doi.org/10.1016/j.phrs.2012.11.007

Roskoski, R. (2016). Classification of small molecule protein kinase inhibitors based upon the structures of their drug-enzyme complexes. *Pharmacological Research*, *103*, 26–48. https://doi.org/10.1016/j.phrs.2015.10.021

Roskoski, R. (2017). Anaplastic lymphoma kinase (ALK) inhibitors in the treatment of ALK-driven lung cancers. *Pharmacological Research*. https://doi.org/10.1016/j.phrs.2017.01.007

Rothenstein, J. M., & Chooback, N. (2018). ALK inhibitors, resistance development, clinical trials. *Current Oncology*, *25*(June), 59. https://doi.org/10.3747/co.25.3760

Sabari, J. K., Santini, F., Schram, A. M., Bergagnini, I., Chen, R., Mrad, C., … Drilon, A. (2017). The activity, safety, and evolving role of brigatinib in patients with ALK-rearranged non-small cell lung cancers. *OncoTargets and Therapy*, *Volume 10*, 1983–1992. https://doi.org/10.2147/OTT.S109295

Sánchez-Torres, J. M., Viteri, S., Molina, M. A., & Rosell, R. (2013). BRAF mutant non-small cell lung cancer and treatment with BRAF inhibitors. *Translational Lung Cancer Research*, *2*(3), 244–250. https://doi.org/10.3978/j.issn.2218-6751.2013.04.01

Seger, R., & Krebs, E. G. (1995). The MAPK signaling cascade. *The FASEB Journal*, *9*(9), 726–735. https://doi.org/10.1096/fasebj.9.9.7601337

Sehgal, K., Peters, M. L. B., VanderLaan, P. A., Rangachari, D., Kobayashi, S. S., & Costa, D. B. (2019). Activity of Brigatinib in the Setting of Alectinib Resistance Mediated by ALK I1171S in ALK-Rearranged Lung Cancer. *Journal of Thoracic Oncology*, *14*(1), e1–e3. https://doi.org/10.1016/j.jtho.2018.06.020

Shaw, A. T., & Engelman, J. A. (2013). ALK in Lung Cancer: Past, Present, and Future. *Journal of Clinical Oncology*, *31*(8), 1105–1111. https://doi.org/10.1200/JCO.2012.44.5353

Shaw, A. T., Friboulet, L., Leshchiner, I., Gainor, J. F., Bergqvist, S., Brooun, A., … Engelman, J. A. (2016). Resensitization to Crizotinib by the Lorlatinib ALK Resistance Mutation L1198F. *New England Journal of Medicine*, *374*(1), 54–61. https://doi.org/10.1056/NEJMoa1508887

Shaw, A. T., Kim, D.-W., Nakagawa, K., Seto, T., Crinó, L., Ahn, M.-J., … Jänne, P. A. (2013). Crizotinib versus Chemotherapy in Advanced ALK -Positive Lung Cancer. *New England Journal of Medicine*, *368*(25), 2385–2394. https://doi.org/10.1056/NEJMoa1214886

Singh, P. K., Singh, H., & Silakari, O. (2016). Kinases inhibitors in lung cancer: From benchside to bedside. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, *1866*(1), 128–140. https://doi.org/10.1016/j.bbcan.2016.07.002

Soejima, K., Yasuda, H., & Hirano, T. (2017). Osimertinib for EGFR T790M mutation-positive non-small cell lung cancer. *Expert Review of Clinical Pharmacology*, *10*(1), 31–38. https://doi.org/10.1080/17512433.2017.1265446

Song, Z., Wang, M., & Zhang, A. (2015). Alectinib: a novel second generation anaplastic lymphoma kinase (ALK) inhibitor for overcoming clinically-acquired resistance. *Acta Pharmaceutica Sinica B*, *5*(1), 34–37. https://doi.org/10.1016/j.apsb.2014.12.007

Spagnolo, F., Ghiorzo, P., & Queirolo, P. (2014). Overcoming resistance to BRAF inhibition in BRAF-

mutated metastatic melanoma. *Oncotarget*, *5*(21). https://doi.org/10.18632/oncotarget.2602

Stewart, E. L., Tan, S. Z., Liu, G., & Tsao, M.-S. (2015). Known and putative mechanisms of resistance to EGFR targeted therapies in NSCLC patients with EGFR mutations-a review. *Translational Lung Cancer Research*, *4*(1), 67–81. https://doi.org/10.3978/j.issn.2218-6751.2014.11.06

Sullivan, I., & Planchard, D. (2016). ALK inhibitors in non-small cell lung cancer: the latest evidence and developments. *Therapeutic Advances in Medical Oncology*, *8*(1), 32–47. https://doi.org/10.1177/1758834015617355

Sullivan, R. J., & Flaherty, K. T. (2013). Resistance to BRAF-targeted therapy in melanoma. *European Journal of Cancer*, *49*(6), 1297–1304. https://doi.org/10.1016/j.ejca.2012.11.019

Takigawa, N. (2018). How Should We Treat Alectinib-Refractory ALK-Positive Non–Small Cell Lung Cancer? *Journal of Thoracic Oncology*, *13*(10), 1438–1440. https://doi.org/10.1016/j.jtho.2018.07.009

Tchekmedyian, N., Ali, S. M., Miller, V. A., & Haura, E. B. (2016). Acquired ALK L1152R Mutation Confers Resistance to Ceritinib and Predicts Response to Alectinib. *Journal of Thoracic Oncology*, *11*(7), e87–e88. https://doi.org/10.1016/j.jtho.2016.03.018

Thomas, A., Rajan, A., & Giaccone, G. (2012). Tyrosine Kinase Inhibitors in Lung Cancer. *Hematology/Oncology Clinics of North America*, *26*(3), 589–605. https://doi.org/10.1016/j.hoc.2012.02.001

Tran, P. N., & Klempner, S. J. (2016). Focus on Alectinib and Competitor Compounds for Second-Line Therapy in ALK-Rearranged NSCLC. *Frontiers in Medicine*, *3*(November), 1–6. https://doi.org/10.3389/fmed.2016.00065

Treiber, D. K., & Shah, N. P. (2013). Ins and Outs of Kinase DFG Motifs. *Chemistry & Biology*, *20*(6), 745–746. https://doi.org/10.1016/j.chembiol.2013.06.001

Trott, O., & Olson, A. J. (2009). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, *8*(8), NA-NA. https://doi.org/10.1002/jcc.21334

Tsai, J., Lee, J. T., Wang, W., Zhang, J., Cho, H., Mamo, S., … Bollag, G. (2008). Discovery of a selective inhibitor of oncogenic B-Raf kinase with potent antimelanoma activity. *Proceedings of the National Academy of Sciences*, *105*(8), 3041–3046. https://doi.org/10.1073/pnas.0711741105

Uhlen, M., Fagerberg, L., Hallstrom, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., … Ponten, F. (2015). Tissue-based map of the human proteome. *Science*, *347*(6220), 1260419–1260419. https://doi.org/10.1126/science.1260419

Umapathy, G., Mendoza-Garcia, P., Hallberg, B., & Palmer, R. H. (2019). Targeting anaplastic lymphoma kinase in neuroblastoma. *APMIS*, *127*(5), 288–302. https://doi.org/10.1111/apm.12940

van Kempen, L. C., Wang, H., Aguirre, M. L., Spatz, A., Kasymjanova, G., Vilacha, J. F., … Small, D. (2018). Afatinib in Osimertinib-Resistant EGFR ex19del/T790M/P794L Mutated NSCLC. *Journal of Thoracic Oncology*, *13*(9), e161–e163. https://doi.org/10.1016/j.jtho.2018.04.020

Vijayalakshmi, R., & Krishnamurthy, A. (2011). Targetable “Driver” Mutations in Non Small Cell Lung Cancer. *Indian Journal of Surgical Oncology*, *2*(3), 178–188. https://doi.org/10.1007/s13193-011-0108-0

Vijayan, R. S. K., He, P., Modi, V., Duong-Ly, K. C., Ma, H., Peterson, J. R., … Levy, R. M. (2015). Conformational Analysis of the DFG-Out Kinase Motif and Biochemical Profiling of Structurally Validated Type II Inhibitors. *Journal of Medicinal Chemistry*, *58*(1), 466–479. https://doi.org/10.1021/jm501603h

Voon, P. J., Tsui, D. W. Y., Rosenfeld, N., & Chin, T. M. (2013). EGFR Exon 20 Insertion A763-Y764insFQEA and Response to Erlotinib—Letter. *Molecular Cancer Therapeutics*, *12*(11), 2614–2615. https://doi.org/10.1158/1535-7163.MCT-13-0192

Vyse, S., & Huang, P. H. (2019). Targeting EGFR exon 20 insertion mutations in non-small cell lung cancer. *Signal Transduction and Targeted Therapy*, *4*(1), 5. https://doi.org/10.1038/s41392-019-0038-9

Wan, P. T. ., Garnett, M. J., Roe, S. M., Lee, S., Niculescu-Duvaz, D., Good, V. M., … Marais, R. (2004). Mechanism of Activation of the RAF-ERK Signaling Pathway by Oncogenic Mutations of B-RAF. *Cell*, *116*(6), 855–867. https://doi.org/10.1016/S0092-8674(04)00215-6

Wang, J., Yao, Z., Jonsson, P., Allen, A. N., Qin, A. C. R., Uddin, S., … Pratilas, C. A. (2018). A Secondary Mutation in BRAF Confers Resistance to RAF Inhibition in a BRAF V600E -Mutant Brain Tumor. *Cancer Discovery*, *8*(9), 1130–1141. https://doi.org/10.1158/2159-8290.CD-17-1263

Wang, S., Cang, S., & Liu, D. (2016). Third-generation inhibitors targeting EGFR T790M mutation in advanced non-small cell lung cancer. *Journal of Hematology & Oncology*, *9*(1), 34. https://doi.org/10.1186/s13045-016-0268-z

Wasserman, H. (2015). Cancer Facts and Statistics. *American Cancer Society*, 19–21. https://doi.org/10.1056/NEJMc1102459

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., … Schwede, T. (2018). SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Research*, *46*(W1), W296–W303. https://doi.org/10.1093/nar/gky427

Whittaker, S., Kirk, R., Hayward, R., Zambon, A., Viros, A., Cantarino, N., … Marais, R. (2010). Gatekeeper Mutations Mediate Resistance to BRAF-Targeted Therapies. *Science Translational Medicine*, *2*(35), 35ra41-35ra41. https://doi.org/10.1126/scitranslmed.3000758

Yang, J. C.-H., Ahn, M.-J., Kim, D.-W., Ramalingam, S. S., Sequist, L. V., Su, W.-C., … Jänne, P. A. (2017). Osimertinib in Pretreated T790M-Positive Advanced Non–Small-Cell Lung Cancer: AURA Study Phase II Extension Component. *Journal of Clinical Oncology*, *35*(12), 1288–1296. https://doi.org/10.1200/JCO.2016.70.3223

Yang, J. C. H., Sequist, L. V., Geater, S. L., Tsai, C.-M., Mok, T. S. K., Schuler, M., … Wu, Y.-L. (2015). Clinical activity of afatinib in patients with advanced non-small-cell lung cancer harbouring uncommon EGFR mutations: a combined post-hoc analysis of LUX-Lung 2, LUX-Lung 3, and LUX-Lung 6. *The Lancet Oncology*, *16*(7), 830–838. https://doi.org/10.1016/S1470-2045(15)00026-1

Yoda, S., Lin, J. J., Lawrence, M. S., Burke, B. J., Friboulet, L., Langenbucher, A., … Shaw, A. T. (2018). Sequential ALK Inhibitors Can Select for Lorlatinib-Resistant Compound ALK Mutations in ALK-Positive Lung Cancer. *Cancer Discovery*, *8*(6), 714–729. https://doi.org/10.1158/2159-8290.CD-17-1256

Yoshida, K., Yatabe, Y., Park, J. Y., Shimizu, J., Horio, Y., Matsuo, K., … Hida, T. (2007). Prospective Validation for Prediction of Gefitinib Sensitivity by Epidermal Growth Factor Receptor Gene Mutation in Patients with Non-Small Cell Lung Cancer. *Journal of Thoracic Oncology*, *2*(1), 22–28. https://doi.org/10.1097/01243894-200701000-00006

Yu, G., Xie, X., Sun, D., Geng, J., Fu, F., Zhang, L., & Wang, H. (2015). EGFR mutation L747P led to gefitinib resistance and accelerated liver metastases in a Chinese patient with lung adenocarcinoma. *International Journal of Clinical and Experimental Pathology*, *8*(7), 8603–8606.

Yu, X., Sharma, K. D., Takahashi, T., Iwamoto, R., & Mekada, E. (2002). Ligand-independent Dimer Formation of Epidermal Growth Factor Receptor (EGFR) Is a Step Separable from Ligand-induced EGFR Signaling. *Molecular Biology of the Cell*, *13*(7), 2547–2557. https://doi.org/10.1091/mbc.01-08-0411

Zeng, Y., & Feldman, A. L. (2016). Genetics of anaplastic large cell lymphoma. *Leukemia & Lymphoma*, *57*(1), 21–27. https://doi.org/10.3109/10428194.2015.1064530

Zhang, H. (2016). Osimertinib making a breakthrough in lung cancer targeted therapy. *OncoTargets and Therapy*, *Volume 9*, 5489–5493. https://doi.org/10.2147/OTT.S114722

Zhang, Q., Zhang, X.-C., Yang, J.-J., Yang, Z.-F., Bai, Y., Su, J., … Wu, Y.-L. (2018). EGFR L792H and G796R: Two Novel Mutations Mediating Resistance to the Third-Generation EGFR Tyrosine Kinase Inhibitor Osimertinib. *Journal of Thoracic Oncology*, *13*(9), 1415–1421. https://doi.org/10.1016/j.jtho.2018.05.024

ZHANG, W., & LIU, H. T. (2002). MAPK signal pathways in the regulation of cell proliferation in mammalian cells. *Cell Research*, *12*(1), 9–18. https://doi.org/10.1038/sj.cr.7290105

Supplementary figures

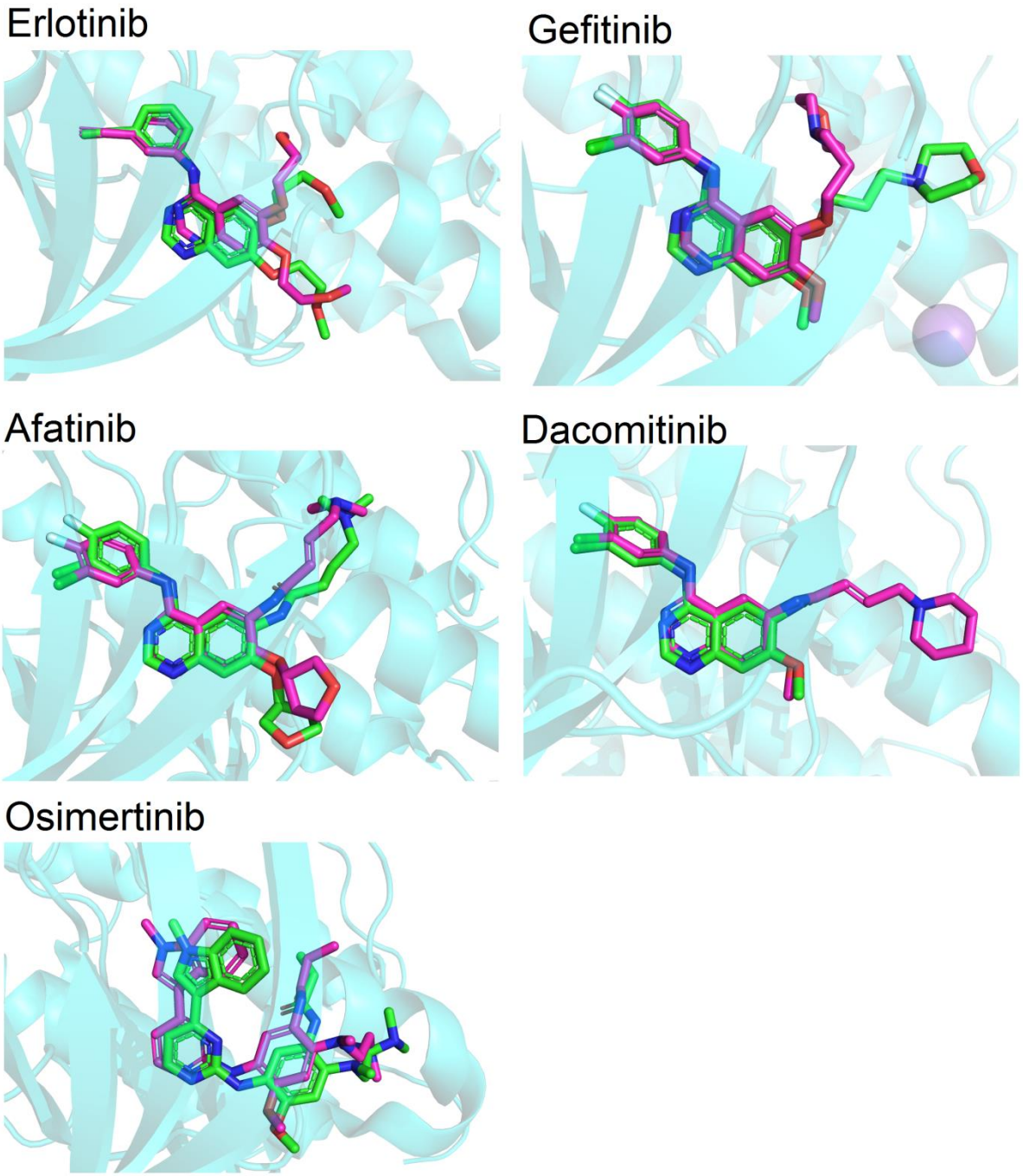


Figure S 1 – Crossdocking binding poses representative of those passing the RMSD filter for EGFR. The crystal structure binding mode for each drug is shown in green for reference. The generated pose is shown in magenta. Nitrogen, Oxygen, Chlorine and Fuorine are colored blue, red, green and white respectively.

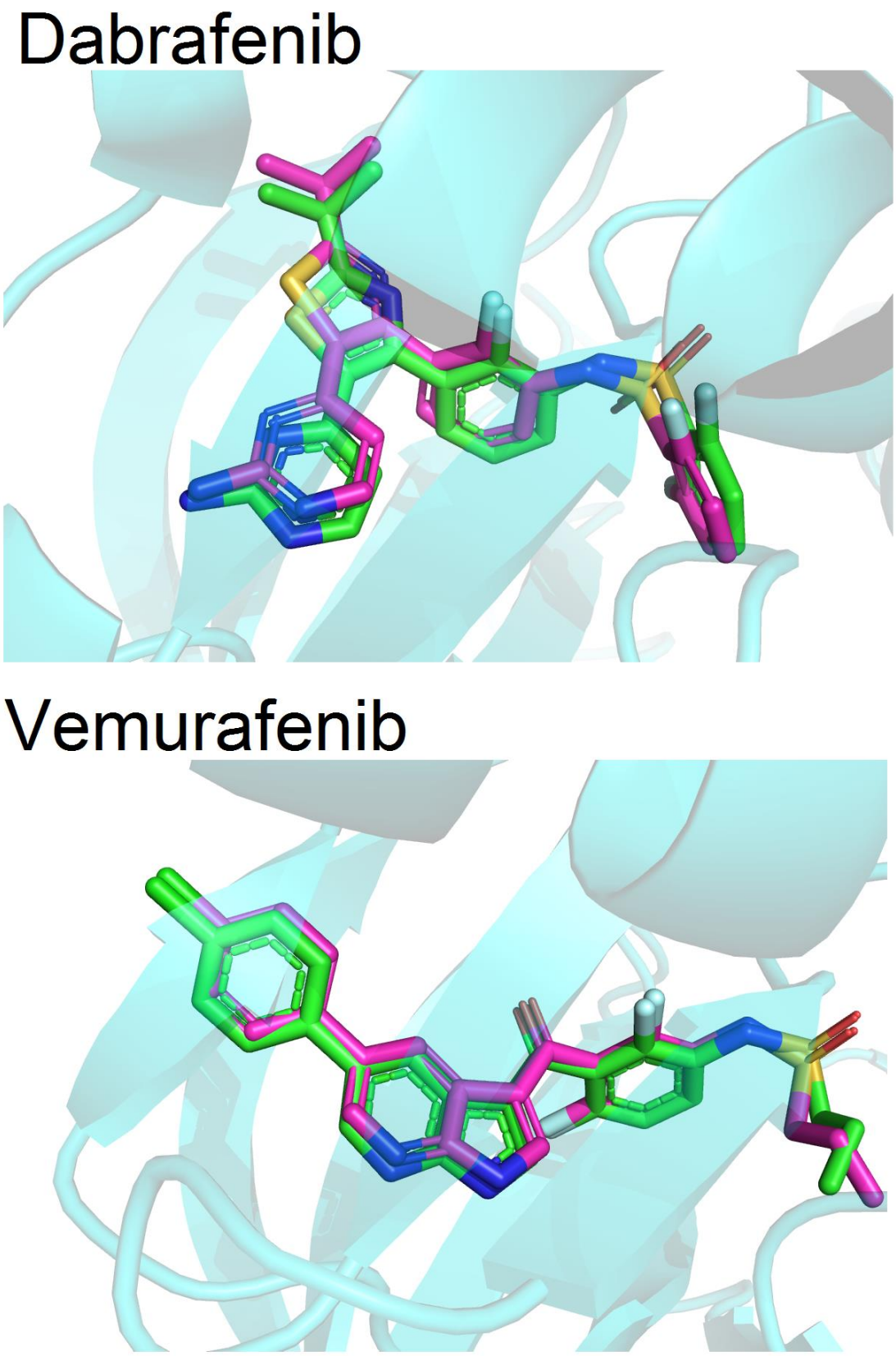


Figure S 2 - Crossdocking binding poses representative of those passing the RMSD filter for EGFR. The crystal structure binding mode for each drug is shown in green for reference. The generated pose is shown in magenta. Nitrogen, Oxygen, Fuorine and Sulfur are colored blue, red, white and orange respectively.

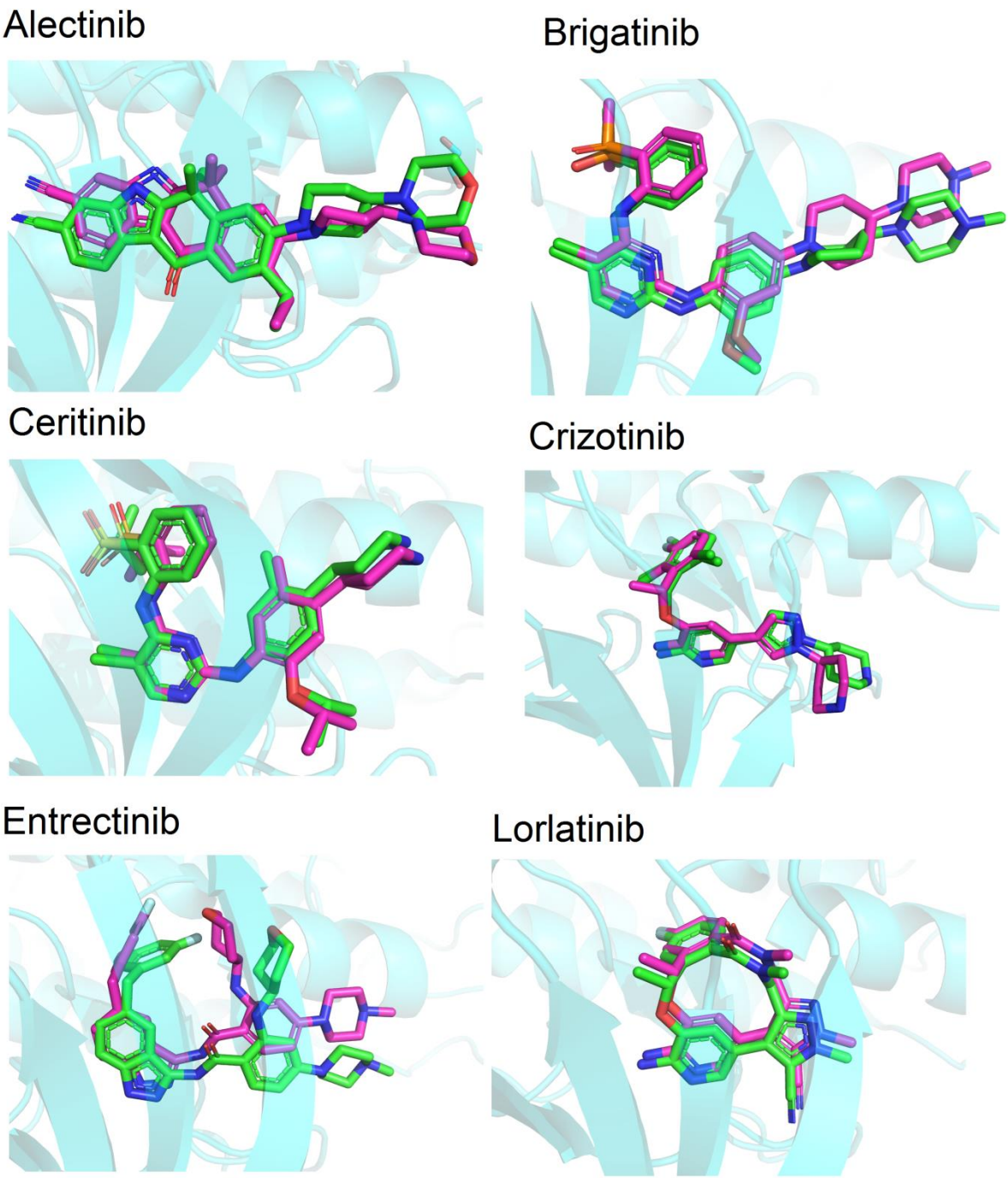


Figure S 3 - Crossdocking binding poses representative of those passing the RMSD filter for ALK. The crystal structure binding mode for each drug is shown in green for reference. The generated pose is shown in magenta. Nitrogen, Oxygen, Fuorine, Chlorine and Sulfur are colored blue, red, white, green and orange respectively.

Table S 1 - Normalized Hydrogen bond strength for hydrogen bonds with hinge region residues in PDB structures of EGFR, ALK and BRAF in complex with kinase inhibitor.

PDB code	H-bond strength (normalized)
4g5j	-.2047

4g5p	-.2255
3aox	-.8714
5j7h	-.5274
4mkc	-.4884
2wgj	-.6405
2xp2	-.6284
2yfx	-.5676
3zbf	-.4390
4anq	-.6252
4ans	-.5706
5aaa	-.5951
5aab	-.4996
5aac	-.4930
4xv2	-.6315
5csw	-.7295
5hie	-.6514
4i23	-.8594
4i24	-.6892
5fto	-.5387
5ftq	-.6785
5kvt	-.4457
1m17	-.9042
4hjo	-.6378
2ito	-.4622
2ity	-.9247
2itz	-.9538
3ug2	-.7582
4i22	-.7697
4wkq	-.5605
5y7z	-.5193
5y80	-.4872
1xkk	-.6284
4cli	-.7827
4clj	-.6523
5a9u	-.6180
5aa8	-.5023
5aa9	-.5479
2jiv	-.3113
3w2q	-.4470
4zau	-.4982
1uwh	-.4621
1uwj	-.1944
3heg	-.1509
3rgf	-.2705

3wze	-.2888
4asd	-.3650
5hi2	-.3005
2ivu	-.4548
3og7	-.4844
4rzv	-.2400
5hes	-.7030