Norges miljø- og
biovitenskapelige
universitet

Master's Thesis 2017 60 ECTS
Faculty of Chemistry, Biotechnology and Food Science

Using AutoDock Vina to dock the proteins TGF-α and EGF to oncoprotein Her2.

Jonas Svåsand
Biotechnology

Acknowledgements

The work of this thesis was conducted at the Faculty of Chemistry, Biotechnology and Food Science, at the Norwegian University of Life Sciences (NMBU).

First and foremost I want to thank my advisor, Professor Lars Skjeldal, for giving me a thesis in a very interesting field. Your enthusiasm for this project coupled with your never ending positivity and encouragement has been an enormous help.

I also want to thank Professor dr. med. Henrik Huitfeldt at the University of Oslo/Rikshospitalet, for the chance to work with this very interesting subject, and also helping me with understanding the Her signaling system.

Ås, Mai 2017

Jonas Svåsand

Sammendrag

Her2 er et transmembrant reseptor protein, som er en del av det cellulære signaliseringssystemet ansvarlig for blant annet celledeling. Her2 er kjent for å være overekspressert i flere forskjellige typer kreft og er delaktig i utviklingen av sykdommen. Selv om Her2 er et reseptor protein, har det ingen kjente ligander. Resultater fra et forsøk med Her2 og de to Her1 ligandene TGF-α og EGF, indikerte at en eller begge av disse hadde mulighet for binding til Her2. Det ble derfor satt i gang forsøk hvor bindingsmulighetene mellom Her2 og TGF-α/EGF ble utforsket via *In Silico* metoden Molecular Docking. Programmet AutoDock Vina ble brukt til å utføre molecular docking og mange områder av Her2 proteinet ble gjennomsøkt for binding til de to ligandene. Det ble funnet flere bindinger mellom både TGF-α og Her2 samt EGF og Her2. De resulterende strukturene av de to ligandene docket til Her2 ble videre undersøkt. Basert på strukturene fra AutoDock Vina og annen data fra docking programmet, ble det funnet at binding mellom TGF-α/EGF og Her2 er en definitiv mulighet.

Abstract

The receptor protein Her2 is involved in signaling important for cell proliferation and is also known for its involvement in several types of cancer. Even though Her2 is considered a Receptor Tyrosine Kinase, there are currently no known ligands for Her2. Results from experiments involving Her2 and known Her1 ligands TGF-α and EGF, indicates that a binding may take place between one or both of the ligands and Her2. To investigate this, there were done several *In Silico* docking experiments between the two ligands and the receptor protein. The docking experiments were done using the molecular docking program AutoDock Vina. The docking experiments were successful and there were found four successful bindings, between Her2 and TGF-α/EGF. All of the different docked structures were evaluated based on the binding affinity and the structures themselves. Based on the different results from the Molecular Docking experiments there can be concluded that there are definitive possibilities of bindings between Her2 and TGF-α or EGF.

Contents

Glossary

Binding affinity – The free energy of the binding, calculated by the docking software

DALI – A protein structure alignment tool/server

Dimerization arm – The part of the Her ectodomains responsible for interactions and bindings between the two ectodomains in a dimer. Located in domain II

Ectodomain – The extracellular part/domains of a transmembrane protein

EGFR – Alternative name for Her1

ErbB family - alternative name for Her family

Fab – Two light chains bound to the two most outermost heavy chain domains of an antibody

Gridbox – The terminology used by the AutoDock programs for a search area

Her2 - A RTK protein, also known by the name ErbB2 and Neu (specifically Her2 from rats)

Herceptin – Brand name for cancer medicine and her2 dimerization blocker Trastuzumab

IgG – Immunoglobulin G, an antibody consisting of two light chains and two heavy chains

Pdb – protein database, also the file extension that is used for making protein structures

Perjeta – Brand name for cancer medicine and her2 dimerization blocker Pertuzumab

RTK – Receptor Tyrosine Kinase, a type of transmembrane receptor proteins.

PDB structures:

1EGF – Structure of EGF (from mouse)

1IVO – Structure of a Her1 ectodomain dimer with EGF bound as ligand

1MOX – Structure of a Her1 ectodomain dimer with TGF-α bound as ligand

1N8Y – Structure of a the ectodomain of Neu (Her2 from rat)

1N8Z – Structure of the ectodomain of Her2 in complex with a Herceptin Fab

2A91 – Structure of domains I-III of the Her2 ectodomain.

2TGF – Structure of TGF-α

3WLW – Structure of a Her2 ectodomain dimer stabilized with two IgG fabs.

The references of the different structures are given the first time they're mentioned in the thesis.

Aim

The aim of this thesis is to use molecular docking software to determine if the oncoprotein Her2 binds the ligand proteins TGF-α and EGF.

Introduction

There is a great search for molecules that can bind to proteins of interest. This search originates from a large need for new medicines, and the quest to further understand the biological mechanisms of the living organisms. By searching for bindings and interactions, there is hope that inhibitors to inhibit disease-causing proteins or inhibitors for oncoproteins will be found. Bindings that enhance the action of essential enzymes and proteins, which could become revolutionary medicines is another area of interest and research. There is also hope that it could lead to further understanding of the binding sites and interactions in the cell, and thereby increase the knowledge about the mechanisms of life. Finding these bindings through Molecular Docking is a method that can provide large advances in the fields of medicine, biochemistry and molecular biology where traditional methods of finding bindings and their structures are time consuming and expensive.

Molecular docking is a method that lately has become even more popular as more protein structures have become available and computer power more powerful. In Molecular docking, the objective is to try and predict the binding and interactions between two different molecules. Usually one of them is a protein and the other a ligand of some sort (including proteins). There are a plethora of different docking programs available today [1], utilizing several different methods and approaches to how these methods are executed. One of these programs is AutoDock Vina.

Molecular docking is not a field without its troubles or roadblocks. Our current models and knowledge are not perfect for predicting how molecules actually interact and we have to either use approximations (based on our current knowledge) of how molecules interact or use other methods, such as calculating the potential energy of the bonds and interactions. The fact that no program is universal is also a problem; they each have their strength and weaknesses. The different programs are also usually specialized for either protein-nonprotein docking or protein-protein docking[1].

Even though these various docking programs and methods are not perfect they are good enough to provide models and answers which can lead to scientific breakthroughs. Docking is becoming widely accepted as a good method for drug discovery and there have been made several discoveries by using molecular docking [2-4]. With molecular docking it is possible to check for possible bindings between molecules and proteins without doing difficult or/and expensive lab work. With a healthy amount of skepticism of the results and a modest and humble use of the docking software combined with respect for its limitations, docking will be able to produce good results. If docking really reaches its full potential it could help with discovering unknown complexes and if e.g. protein-protein docking is improved it could help with providing structures for many unknown complexes, just for the human interactome (interactions of proteins etc.) there is great potential for increasing our knowledge in the field, as it is estimated only 8% of interactions are known. [5]

The docking software used in this thesis is AutoDock Vina**,** a program made by Oleg Trott of The Scripps Research Institute[6]. This program requires manual input in the form of a search area of the receptor, which represents the area the docking of the ligand will be attempted on the receptor. In order to choose a search area that will have the chance to yield results while not taking too much time, requires much thought and research. While getting to know the software is time consuming, the real skill in using the AutoDock Vina software is in the choice of the search area.

The oncoprotein Her2 is a receptor protein without any known ligands. In order to check for ligands numerous experiments and searches have been done but no ligands have been found[7]. In an experiment on the Her2 signaling system conducted by Professor Henrik Huitfeldt at the University of Oslo, the results indicated that TGF-α and EGF might bind to Her2. Epidermal Growth Factor (EGF) and Transforming Growth Factor Alpha (TGF-α) are known ligands to the Epidermal Growth Factor Receptor (EGFR, also known as Her1) which belongs to the same family as Her2. Conducting laboratory experiments to further check for the binding of these two ligands to Her2 is expensive and time consuming. In order to not run the experiments in vain and waste resources and time, a docking experiment was suggested to check for a binding between these ligands and Her2.

This thesis will explore the possibility of binding, for the two proteins TGF-α and EGF to the oncoprotein Her2. The experiments to check for binding will be conducted entirely *In Silico*, and the results from the *In Silico* docking experiments will be used to discuss whether there is a possibility for Her2 to bind one or both of these ligands.

Theory and Methods:

The ErbB/Her family

The Her family, a family of receptors responsible for cell growth and proliferation

The Her signaling system is made up of four proteins, Her1, Her2, Her3 and Her4 also known as ErbB1,ErbB2, ErbB3 and ErbB4 respectively, which are all Receptor Tyrosine Kinases (RTK). The Her family are paralogs that were created by gene duplication and mutation[8]. The RTKs are transmembrane proteins, consisting of four extracellular domains (conjointly known as the "ectodomain") that bind ligands. Linked to the end of the ectodomain is a transmembrane domain and a juxtamembrane domain that together serves as a link between the ectodomain and intracellular domains. The intracellular domains have kinase activity, and phosphorylation sites that serves as binding sites for intracellular signaling proteins. The ErbB/Her family are extracellular receptors, which bind ligands that trigger the formation of a dimer of the ectodomains. These dimers can either consist of two of the same Her proteins (homodimer) or two different members of the family forming a dimer (heterodimer), see figure 1A for the most common pairings. After formation of the dimer, a signaling system will be triggered.

The ErbB family is involved in different pathways involved in cell growth and proliferation, and is critical in development and growth of Humans. Her1 is maybe the most important; it binds ligands such as the Epidermal Growth Factor (EGF) and Transforming Growth Factor Alpha (TGF-α). The binding of the ligands is responsible for the dimerization and subsequent autophosphorylation of the kinase domains. The phosphorylated kinases then recruits and phosphorylates other proteins to trigger intracellular pathways. Amongst the recruited proteins are proteins such as Grb2 and Sos, with Sos being able to recruit and activate the G protein Ras, which in turn activates the MAP cascade, initiating cell proliferation. Other pathways activated by the ErbB family includes the Akt pathway and the activation of Phospolipase Cγ [9]. The ErbB family is a very important family of proteins due to its participation in signaling and triggering pathways important in growth and thus also cancer. The importance can be seen in trials where the *erbB-1* gene of mice was mutated, the mutation resulted in either severe deformities or death for the mice [10]. There can also be seen deformities in mice where Her2 and Her3 being knocked out (they being dependent on each other as Her2 can't bind a ligand and Her3 don't have kinase activity and therefore can't form a functional homodimer) which shows the importance of the Her family for survival.

The start of the signaling cascade is the ligand binding to the ectodomain and the subsequent dimerization of the ectodomains of two Her family members (shown in figure 1A on the next page). This leads to the autophosphorylation of the intracellular kinase domains and subsequent recruitment and phosphorylation of the other proteins. When the Her family

member receptors (except Her2) are free of bound ligands and non-dimerized they are found in an inactive form that has no possibilities of dimerizing until they are activated by a ligand.

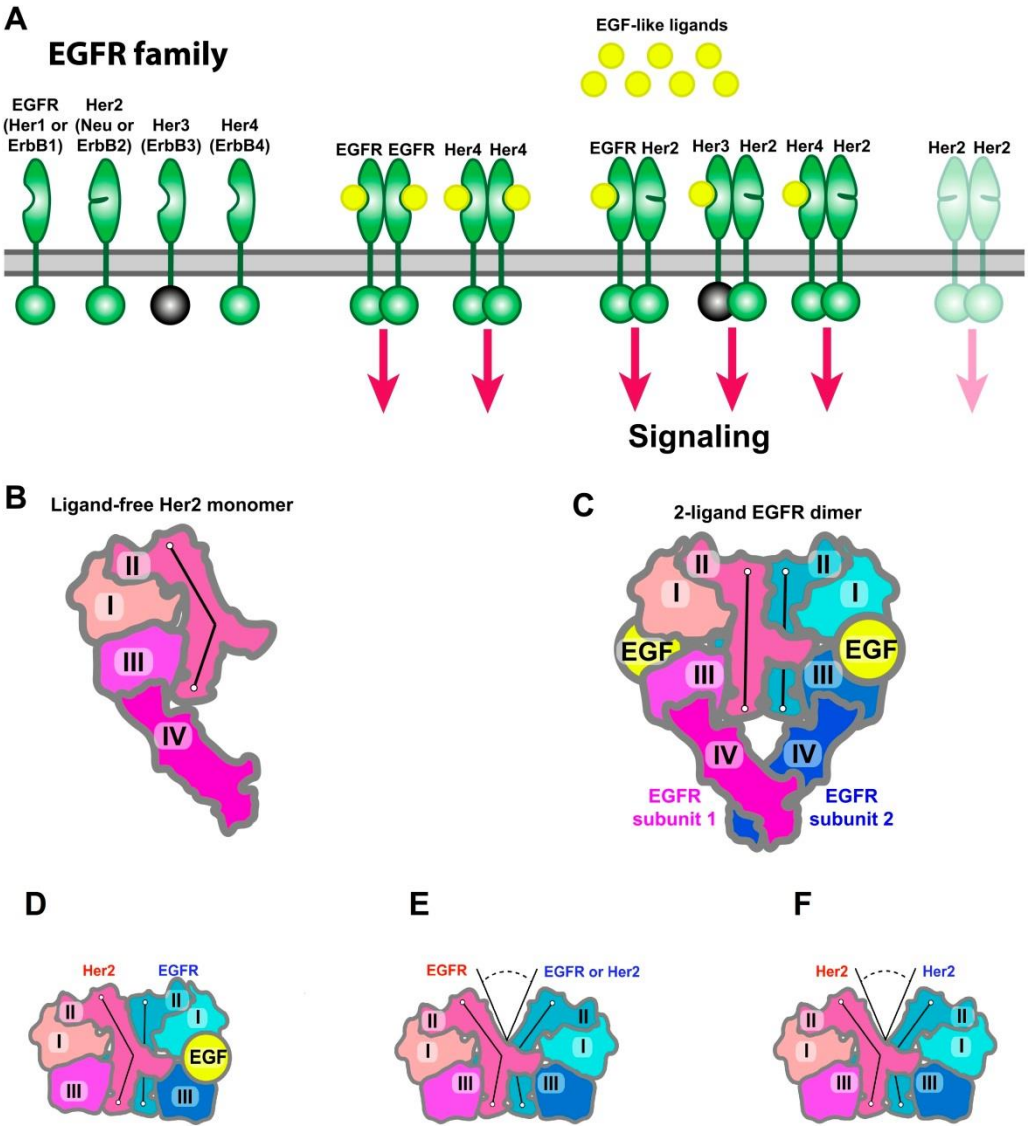

Figure 1: A) Showing, in general terms, how the binding of ligands and dimerization work for Her family members, as well as which pairs are functional dimer partners. It also shows that when two ectodomains dimerize, the intracellular domains will be brought together, and they will phosphorylate each other (phosphorylation not shown). Her2s ectodomain is closed due to not binding ligands, and Her3s intracellular domain is black, as Her3 does not have kinase activity. **B)** The ectodomain of a Her2 monomer. The bend in domain II is included. **C)** A EGFR (Her1) ectodomain dimer, made up of two EGFR proteins, each bound to a EGF molecule. It shows the interaction between the two dimerization arms in domain II, and that domain II is straight with EGF bound. **D)** A dimer between Her2 and a EGFR (with EGF as ligand). The dimer is possible despite Her2s domain II being bent. **E)** A dimer between two EGFR (Her1) ectodomains without any ligands bound. The lack of ligand leaves domain II in a bent configuration. The dimer is unstable/unfavorable because of the conformation of domain II, and this is shown by the large gap between the top parts of the two domain IIs. **F)** A dimer between two Her2 ectodomains, the structure of domain II makes the dimer unstable/unfavorable. This is shown by the large gap between the top parts of the two domain IIs
Domain IV has not been included in figures D),E) and F). As domain IV is not involved in dimerization this exclusion should not matter. These three figures are the results of simulations of EGFR and Her2.
Figure taken from article by Arkhipov et al. (2013) [11]
The figure is a combination of figures 1 and 4 from the aforementioned article, and was edited to suit this thesis.

The part of the ectodomain that is responsible for the binding to another Her protein is the dimerization arm (residues 245-266 in Her2)[12] found in domain II. In the inactive form of the Her proteins, the dimerization arm is bound to domain IV of the ectodomain[13, 14] (see figure 2A). In this conformation the dimerization arm has no possibility to form bonds with another monomer due to it already being bound to the domain IV, which acts as a steric blocker. Her2 differs from the other Her family members in that its dimerization arm does not bind to domain IV, and is therefore always in the active form[15]. The dimerization cannot happen without a change in the structure of the domains and how they are positioned in relation to each other. When the Her family ligands bind to the ectodomains they trigger a rearrangement of the domains by binding to domain I and III, thus bringing them closer to each other. At the same time this change in structure releases the dimerization arm from its binding with domain IV making the dimerization arm ready for a binding[16, 17]. This change in the ectodomain, when binding a ligand is depicted in figure 2.

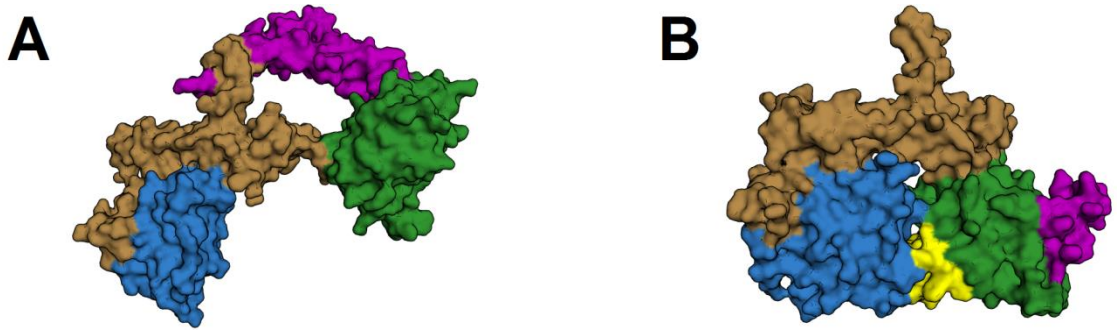

Figure 2: A) A structure of the ectodomain of Her3(pdb: 1M6B[14]) in the inactive conformation, where the dimerization arm is tethered to domain IV. **B)** A structure of the ectodomain of Her1 (pdb: 1IVO) with EGF bound. Her1 is therefore the active conformation and ready for dimerization. The truncation of domain IV in figure B) is a feature of the structure, and not a consequence of releasing the bond to the tether. In both figures domain I is colored blue, domain II colored brown, domain III colored green, domain IV colored purple and EGF colored yellow. Image created in PyMOL[18]

The change induced by the dimerization arm being released from its binding with domain IV makes the dimerization arm and rest of domain II available for binding. The change in structure is caused by the binding of a ligand in the pocket between domain I and III. This ligand binding also changes the structure of the domain II of the Her1/Her3/Her4 proteins. The changes in the structure of domain II consists of making the structure more "straight" and eliminate the slight curve the structure has in domain II, without a bound ligand. This change is seen when you compare the Her1s in figure 1C that are straight to the curved Her1s shown in figure 1E. The straightening of domain II is important for the proteins ability to dimerize, as the curved structure has a structure that is less ideal for dimerization[11], this is illustrated by figure 1E. Her2 has an inherent permanent curve in domain II as shown in figure 1B. It is implied that having a curved domain II for a Her protein, does not impact its ability to form heterodimers, as Her2 do have the ability to form heterodimers. This is reinforced from the molecular dynamics experiments by Arkhipov et al. (2013), that shows that Her2s curved domain II does not make it impossible to form a heterodimer (depicted in figure 1D). While a Her1/Her3/Her4 heterodimer with Her2 is possible a Her2 homodimer is

less stable than the Her2 heterodimer as having two curved domain IIs forming a bond is unfavorable (see figure 1F) and Her2 homodimers is not possible in normal conditions.

To sum up, the binding of a ligand to Her1/Her3/Her4 releases the dimerization arm from a binding with domain IV. At the same time the ligand causes a change in the structure of domain II. The Her family member with a ligand bound is then ready to dimerize with another Her family member that has a ligand bound or the ligand independent Her2 protein. Dimerization of a Her protein happens by the dimerization arm interacting and forming bonds with the dimerization arm of another Her family protein and form a back to back dimer. This leads to activation of the intracellular kinase domains, which will phosphorylate each other and other molecules to trigger cell signaling pathways.

There are many ligands that bind to the Her family, these are usually separated into two groups, EGF agonists which bind and activate EGFR/Her1 of which there are at least 7(some can also bind Her4), and neuregulins which binds Her3 and Her4 [19]. As mentioned above none of the ligands that bind to any of the other three members of the Her family are believed to bind to Her2. In the EGF agonist group, the ligands EGF and TGF-α are found. These two ligands are well known in regards to Her1 and structure for the complex between both EGF and Her1 and TGF-α and Her1 has been resolved, as well as the structure of how these complexes look in a homodimer. The Her ligands with the highest affinity share similar structure that consists of approximately 60 residues, the "EGF-like core domain". The domain is recognizable from the six cysteine residues that are spaced out at a similar rate and form three disulfide bridges[20].

While TGF- α and EGF shares great similarity in structure they are very different when it comes to their functions. TGF- α is first expressed as a pro-protein, pro- TGF- α, which is a protein that is found integrated into the cell membrane with a weight of between 20 and 22 kDa [21]. The pro- TGF- α is then released from the membrane and further truncated in a twostep process, to give the mature TGF- α protein weighing 6 kDa (and being 50 residues long). EGF is also expressed first as a pro-protein that can be found in the cell membrane, before being processed into a 53 residues long mature EGF protein[22]. The difference in function between TGF-α and EGF, is largely in their target. TGF-α is mainly an autocrine ligand, while EGF is mainly endocrine[23].

Her2 an oncoprotein and a Receptor Tyrosine Kinase without receptors

No ligand has been identified for Her2. Its function is believed to be a dimerization partner for other Her family members that has bound a ligand. The conformation of Her1 with EGF bound [8] being similar to Her2s native conformation (a free dimerization arm), implies that Her2 does not need to bind a ligand to dimerize. As Her2 does not homo dimerize in normal conditions [24] and has a conformation that is ready to dimerize, there seems to be no need for binding of a ligand in order to dimerize.

Her2 is especially interesting due to its links to development of cancer. Overexpression of Her2 is found in different type of cancers, but is most prevalent and researched in breast cancer where it is found in between 15% and 30% of cases [25-27]. When expression is amplified the Her2 will form homo and heterodimers more easily, and their signaling is altered [26]. The Her2 homodimer is believed to be less stable and less of a favorable dimer (due to domain II's curve[11])than the rest of the activated Her family's homo and heterodimers [11]. Due to the extreme amount of Her2 found on the cellular surface of these cancer cells, homodimers will still be formed in large number due to the abundance of Her2 ectodomains being ready. Therefore the less stable dimer does not matter as new dimers will form and be broken in large numbers. The overexpression of Her2 is also linked to lower survivability and difference in response to normal chemotherapy treatment. This dimerization is ligand independent, which means that the dimerization can happen without the need of ligands to bind. Because of this there is no need for more (i.e. overexpression of) ligands, like Her1 needs, but just overexpression of Her2 which then results in cell transformation due to the ligand independent Her2 dimerization[19]. Because of this, blocking of HER2-dimerization is interesting in concern to cancer treatment and has spawned treatments for cancers where overexpression of Her2 is present. Two of them being the Her2 dimerization blockers, Herceptin/Trastuzumab and Perjeta/Pertuzumab [27-29].

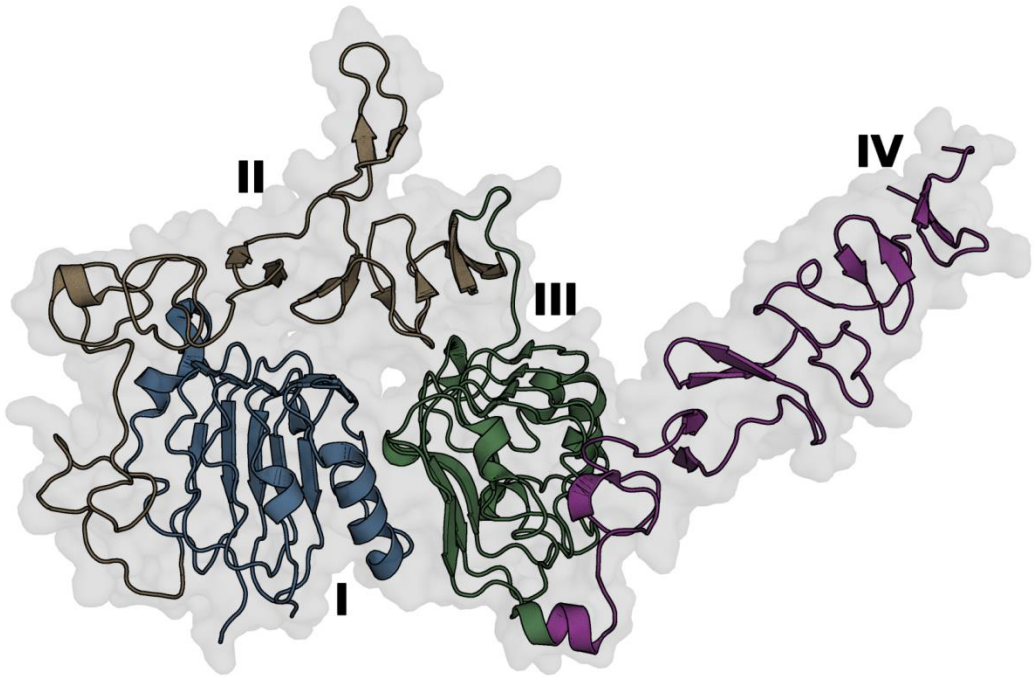

Figure 3: Visualization of the Her2 ectodomain from the pdb structure 1N8Zc. The extracellular domains, I-IV are depicted in different colors. Domain I blue, domain II beige, domain III green and domain IV purple. The grey outline in the figure represents the surface of the protein. Image created in PyMOL [18]

The ectodomain of the Her proteins (the extracellular part of the protein which is involved in ligand binding and dimerization), is usually divided into 4 domains. These are either known by the chronological numbering I, II, III and IV with I starting at the N-terminal, or L1, CR1, L2 and CR2 respectively. The two L domains are very similar and is dubbed as L domains as

they are large (though the L has also later been used to represent the fact the domains are Leucine rich), the two CR domains aren't as similar as the L domains but have been classified as the same domain type since both are Cysteine Rich (CR).In Her1 these domains are defined as follow, domain I/L1 residues 1-165, domain II/CR1 residues 166-310, domain III/L2 residues 311-480 and domain IV/CR2 residues 481-620 [19, 30]. In this thesis the same definition will be used for Her2, i.e. the same residues will represent the different domains. This is illustrated in figure 3 using the Her2 structure 1N8Zc[28] (note that domain IV is truncated, and therefore the structure only contains residues 1-607).

Proteins, their structure, properties and terminology concerning them

A protein is made up of several (usually several hundred) amino acids that have been linked together by peptide bonds. They will form a long peptide chain, called a polypeptide. In one end you find an amino acid, or residue, with an exposed NH group, while on the other end you find an exposed COOH group, these are called the N-terminal and C-terminal respectively.

The 3 dimensional structures of proteins are determined by the interactions and bindings between the different residues in polypeptide chain. These bindings and interactions make it so the almost 2-dimensional polypeptide chain, will create a structure that occupies more of the room. The most important bindings that make up the protein are alfa-helixes and beta-sheets. These are what are called the secondary structure of proteins; the sequence of the polypeptide chain is the primary structure. The alfa helix is a structure made up by hydrogen bonds between residues next to each other in the polypeptide (the NH and CO groups of the main chain peptide bonds, are the ones involved). The β-sheets form hydrogen bonds with more remote residues; these residues have been positioned parallel in relation to each other [31].

There is a lot of terminology used to describe protein structure, the main terms used are topology, architecture and domains[31]. Topology describes the relation in the sequence between the secondary structures, while architecture is the relation of secondary structures in the tertiary structure/3dimentional space. A domain of a protein can be defined as an easily recognized subsection that has a certain property, either its architecture or its amino acid makeup. Proteins can usually be divided into several parts, where each part of the protein is a domain. These domains are something that is also observed throughout the evolution of proteins as several proteins can share some of the same domains.

One of the more important types of proteins in the human body is the Antibodies. The antibodies are able to identify foreign elements in the body, and therefore play a crucial role in the immune system. Different types of antibodies have different structure, and the different structures are themselves varied. This is why the antibodies are able to recognize so many different molecules and/or foreign elements such as viruses. The antibodies are made up of

several polypeptide chains of different sizes that binds together to make an antibody. The polypeptide chains consist of different both different number and different types of domains, a chain with two domains are named a light chain, and a heavy chain consists of four or five domains. How many chains that are involved in a single antibody varies, but they come in pairs of two light and two heavy chains, with the two domains of each light chain bound to the two last domains of a heavy chain.

In order to be able to bind many different molecules (as the immune system encounters a vast number of different molecules) there needs to be a great variance in the antibodies, and these do not only come from using different domains to form the polypeptide chains, but also from the variance in the fab region. The fab consists of the two light chain domains and the two heavy chain domains that are bound to the light chain.

The ability of the fab to bind to so many different molecules have been found useful in medicine where the binding of the fab can act as an inhibitor to an enzyme or a receptor, e.g. the anti-cancer drug Herceptin [28]. The Herceptin fab, made up of two heavy chain domains and two light chain domains, can bind Her2 as is shown in figure 7 on page 21. The antibodies are therefore very useful in research and medicine because of the high variability of fab regions, making it possible to find a fab for a specific target, that will bind it with high specificity[31].

File formats, structure databases and visualization tools

The pdb format is a file format that is made to contain the information of a protein structure. The most important information included, is the absolute position of the atoms of the protein (excluding hydrogens). There is also included information about the secondary structure of the protein, i.e. whether an amino acid is involved in an α-helix, β-sheet or a loop. All this information is included in the pdb format together with information about how the structure was obtained, which organism the protein originates from, the sequence of the polypeptide and more. In summation a pdb file contains the structure of a protein, as well as important information about the protein.

All protein structure pdb files, both from published research and yet to be published research, can be found in the protein databank (accessible at www.rcsb.org). In the database all the structures are sorted by name and description. In addition all the pdb files are given a four letter code (consisting of one number followed by three letters or numbers, not case sensitive). This code while not descriptive can be used to quickly look up more information about a structure in the database.

A pdb file does not necessarily only contain the structure of a single protein or a protein may be made up of more than one polypeptide. To differentiate between different

proteins/polypeptide chains the pdb files will designate each of them as a chain. Different proteins or polypeptide chains can therefore be singled out. A way of denoting a specific chain of a pdb structure is to add the letter of the chain after the pdb code. E.g. to show that only chain d of pdb structure 1ABC was used one would denote the structure used as 1ABCd.

In this thesis most protein structures will be referred to by their four letter pdb code in upper case letters with the chain following the four letter/digit code in lower case letters.

PyMOL is a computer program for visualizing protein structures from pdb files[18].
It is an extensive visualization tool and is the program used for the majority of the figures presented in this thesis. PyMOL offers a large number of options for customization of how the molecule and proteins is presented. It is possible to make certain elements, amino acids, chains and domains appear different. PyMOL also has a limited set of editing tools, alignment tools etc., though these are not that good and there are much better options available.

Capturing images from PyMOL can be done with the inbuilt screenshot option, though the resulting screenshot is not of the best quality. There is an option called Ray tracing which will trace the molecule as you are viewing it at that moment. This produces a high quality still shot, which can then be captured, and creates a high quality figure. All PyMOL figures in this thesis are created by using the ray command on a 1920x1080 resolution.

Docking software and procedure:

AutoDock Vina

AutoDock Vina is a computer program for predicting molecular docking[6]. This is done by using a model where noncovalent bonds in addition to hydrophilic and hydrophobic interactions are predicted between the ligand and receptor. By using a scoring function, AutoDock Vina attempts to calculate an approximate of the standard chemical potential of the system. This will give the free energy of the system, if a negative value is observed the binding is thermodynamically favorable.

Vina cycles through configurations of the ligand docked to the receptor, and calculates the score according to the scoring system[32]. This is done by placing the ligand in a set conformation within the search area, and the binding affinity is then calculated based on the distances between atoms of interest in the ligand and receptor. After the binding affinity is calculated the ligand is placed in a new conformation that differs from the first whilst still in the search space, and the binding affinity is calculated once more. This is then repeated for all conformations that Vina has decided on, and the different configurations are then ranked according to their binding affinity. The starting conformation and all other conformations used are decided by a seed, each run of Vina will use a random seed unless one set a specific seed to be used. Overall AutoDock Vina works by making many different conformations for

the ligand receptor complex and calculates the binding affinity for them, and ranks them accordingly.

Each run of Vina is done by a random seed, that defines which configurations are used that run [32]. Because of the random seed the result of one Vina run might differ of the result of another Vina run, even though the molecules used and search space was identical. Therefore there is a chance that one run of Vina might not necessarily give the optimal result; it also means that the results might differ from one run to another.

If the search area chosen is not big enough the ligand will not necessarily be big enough for the ligand to rotate properly. If the search area on the other hand is too big, it can also cause problems, as Vina will require more time to run its calculations. This means that the search areas used when running Vina has to be chosen carefully and with much thought.

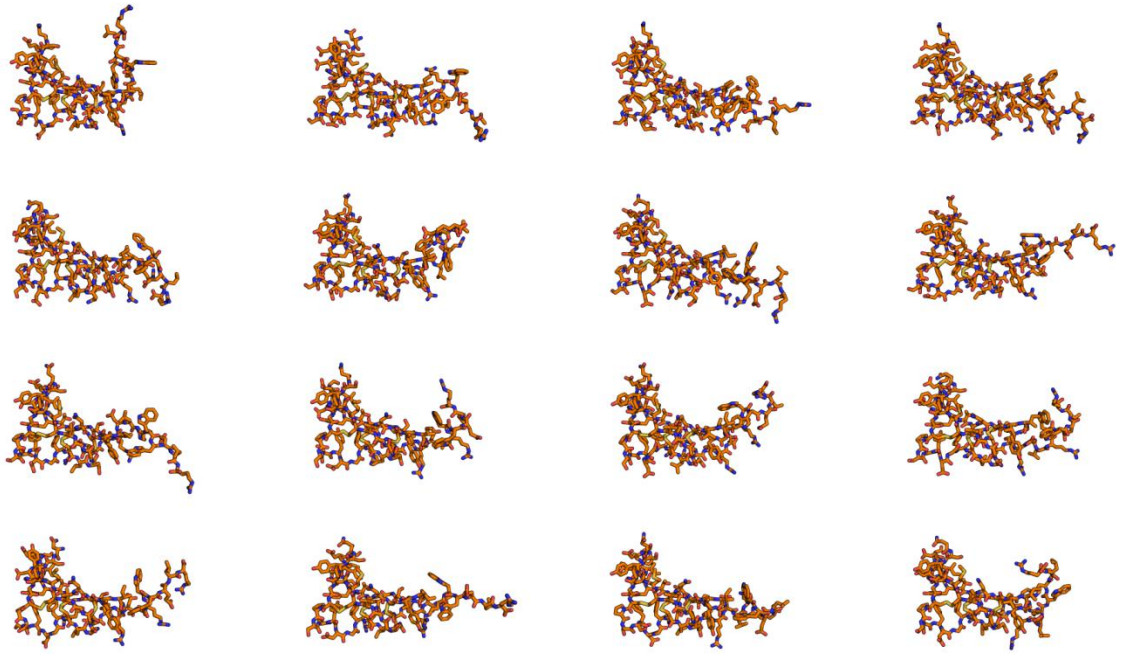

Figure 4: All of the 16 different conformations of EGF. This figure illustrates the internal movement of the protein[31] by showing the 16 different conformations that was found by NMR (pdb 1EGF). Images created in PyMOL [18]

AutoDock Vina also has the option for the ligands and receptors to be flexible. When this option is used during the docking experiment the structures used in the experiment will be flexible. This means that the structures are able to move, this is a useful feature as ligands and receptors can change their structure (not difference in the chemical makeup, but rather that molecules move in relation to each other) and this will then be accounted for. Proteins also tend to move or they are flexible, but this is not reflected in a structure from an x-ray experiment, and Vina can only use one conformation for each structure so this will be the case for NMR structures as well. NMR structures contain at least 10 different conformations of the structure, representing the flexibility of the protein. In the case of pdb 1EGF[33] which is used in this thesis, the structure was determined by NMR and the pdb contains 16 different

conformations shown in figure 4. Using the flex option while running AutoDock Vina, is a method that can be used to try to replicate a proteins natural flexibility in the docking experiment.

A negative effect of allowing the ligand and receptors to be flexible is that they may change into a conformation that has unfavorable torsion angles. For the ligand this can be observed in the output of the docked ligand pdb, and can be checked with a Ramachandran plot (a plot of torsion angles). For the receptor there is no possibility to control how the structure was changed during docking, this can also be a problem. The structure of the receptor may have changed somewhat, and this could cause the docked ligand to not "fit" to the original input receptor structure as it did the flexible receptor during docking. Nevertheless the flexible ligand option does bring more positive than negative as it allows checking for docking in more conformations than the one observed in the available structures in the protein database and is a model that more accurately portrays how proteins actually interact.

Procedure for using AutoDock Vina

For Vina to run, the structure for both the ligand and receptor has to be available in a pdbqt file. The pdbqt structure file format is similar to the pdb file format, but the pdbqt format also includes the polar hydrogens as opposed to pdb files which mostly does not contain information about hydrogens. This is because hydrogens structure/placements are not usually gathered by the modern methods of structure determination. The pdbqt file is made by using the AutoDock Tools program [34] which can calculate the position of the hydrogens from a pdb protein structure and add them and thus creating a pdbqt file.

In addition there is a need for a selection of a search area of the receptor. The search area needs to be specified by size and coordinates in the 3-dimentional space, and Vina will only search for docking possibilities in the specified search area of the receptor. AutoDock Tools has an option called "gridbox" which visualizes the search area, and can be used to assist with the selection of a search area.

Vina was run on a computer network, ABEL[35], where larger amounts of data power could be used for the docking experiments. The docking experiments were run by first converting the ligand and receptor pdb files to pdbqt using AutoDock Tools. Then a standard configuration file for AutoDock Vina was made, containing information about ligand, receptor, search area, computer usage and output information. The docking experiments were then started using the configuration file, note that ABEL is a Linux based computer network and therefore there are some extra steps in executing commands compared to running on a personal Windows computer.

The output of a completed Vina experiments is a modified ligand structure given in a pdbqt file. This file contains the structure of the ligand docked to the receptor. This means that the structure may have been modified compared to the input and of course the coordinates of the

atoms have been changed to their new (docked) position. Vina gives up to 9 results from each docking experiment; all of the structures are given in the one pdbqt file and can be viewed with a visualization program like PyMOL or AutoDock Tools. In addition to the docked structure some data concerning the docking is also given, one set for each of up to 9 results. The output of AutoDock Vina consists of the theoretical binding affinity as well as two measures of how close the results are to each other (two different RMSD measures). The affinity gives some information about how strong the binding is, the smaller the number the better. If the affinity is positive this means that the binding is thermodynamically unfavorable, and that the found binding is most likely a false positive.

Structural comparison, and sequence alignment

In order to better compare protein structures to each other the method of superpositioning is used. This is a way of aligning structures. In superpositioning one of the two proteins is moved around until as many residues as possible are close to/in the same spot as the corresponding residues of the other proteins. It is hard to accommodate all atoms in an attempt to superposition two proteins on each other, and because of this the focus is on trying to align/position the Cα of corresponding residues as close to each other as possible[31]. Other methods can also be applied to make a better structural alignment. One of them is comparing the interactions of residues in the proteins. Throughout evolution, structure is better preserved than the sequence. The interaction between residues must then also be preserved relatively well as the structure is highly dependent on the residue interaction. In essence this method makes a matrix of the distances between the alpha carbons and compares it with a matrix made in the same way from one or more other proteins. This method was implemented into the powerful structural alignment tool DALI(distance-matrix alignment) [36]. DALI is also connected to a database and regularly compares new proteins to all proteins in the database, though this function is only useful if you have a protein of unknown family/use and want to find homologs or proteins with similar function/structure.

Root-mean-square deviation (RMSD) is used as a measurement of the quality of a structural alignment. RMSD is calculated by formula (1).

$$\sqrt{\frac{1}{N}\sum_{i=1}^{N}\delta_i^2} \qquad (1)$$

N is the total number of atoms, numbered 1 to N, and δ_i represents the distance between atoms *i* of the two structures compared, after the superposition has happened. In essence the formula is the root, of the mean of distance squared, which gives you a value that indicates how close the compared atoms are on average. This will also be an indicator of how closely the structure as a whole resembles each other after the superposition[31].

RMSD is also used to describe the similarity between the output structures from AutoDock Vina. The comparison is done between the best match (the one with most negative binding

affinity), and each of the other matches individually. AutoDock Vina gives two different measurements of RMSD, one called “RMSD lower bound” and the other “RMSD upper bound”[32]. The RMSD upper bound uses the same atoms (in the structure) in the different structures as pairs for calculating distance, e.g. it uses the Cα of residue 33 in both structures as a pair etc. The RMSD lower bound is determined by formula (2).

$$rmsd/lb\ (c1, c2)=max(rmsd'(c1, c2), rmsd'\ (c2, c1)) \qquad (2)$$

rmsd’ is a RMSD calculation which uses the two closest atoms of the same elements (in the different structures) as a pair, but as it is an asymmetric RMSD it cannot be used directly and therefore the RMSD lower bound has to be deduced from the rmsd’ to have a symmetrical RMSD.

Another measure for how good a structural comparison is the Z score**.** The Z score is the standard deviation away from the mean, and is a statistical term. I.e. this value is the value you have, with the mean being detracted from it and then the result is divided on the standard deviation. [37]**.** The Z value says something about the grade of deviation in the distance placement of the atoms. It is comparable to a value from a t-test or the p-value for the t-test. The higher Z is the more chance there is that the structural alignment is significant [38]. A Z score of 4 is considered significant [38], while the DALI server warns that alignments with Z-score of less than 2 is not to be trusted.

Measures of structure quality

There are several methods that are used to determine the quality of a structure. Some of these methods tell you how accurate the structure is, whereas others tell about the correctness of the structure. Different types of information is used in different ways some information is used to better the structure, while others cannot improve upon the structure but only tell something about its attributes.

One of the more common methods to determine how the quality of a protein structure is the Sasisekharan-Ramakrishnan-Ramachandran plot(from here on called “Ramachandran plot”) [31] . The Ramachandran plot is a plot of all the torsion angles in the protein, plotted versus the sterically “allowed” angles. The possible or allowed torsion angles are determined by different chemical restrictions. The psi and phi angles are restricted by the best energetically conformation from their side chains, while the omega angle is almost always locked at 180°.

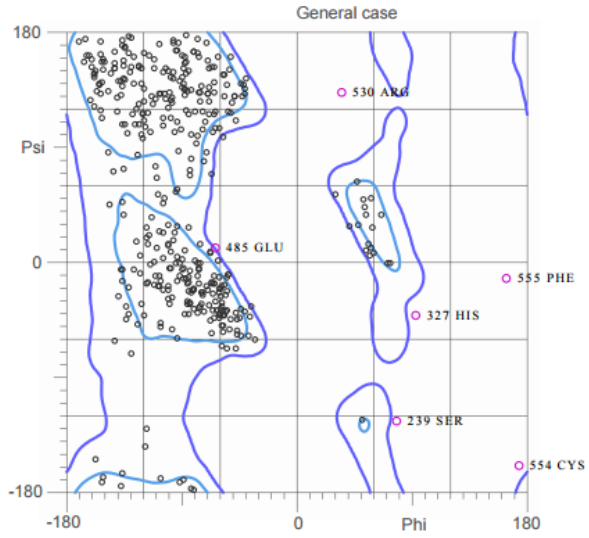

Figure 5: A Sasisekharan-Ramakrishnan-Ramachandran plot of 3WLWa [39] The small circles represents residues. Residues inside a blue area is considered to have a favorable torsion angle, while if a residue is outside the blue area but inside the purple it's considered to have an allowed torsion angle. Residues outside the purple area have unfavorable torsion angles.

In the Ramachandran plot there are two main "allowed" regions that covers the torsion angles that are found in residues involved in β-sheets and α-helixes. In addition there exist some other regions that are allowed such as for glycins that are part in a left handed alpha helix. These three regions can all be observed in figure 5 where the different regions and different residues can be observed. 97, 8% of the residues in the structure 3WLWa are within allowed regions. The Ramachandran plot shows you whether or not the torsion angles in a protein structure are acceptable.

Consideration of which protein structure(s) to use in the docking experiment

The pdb structure 3WLW [12] is only of the extracellular domains of Her2, the structure does not include the part of the extracellular domains that connects to the rest of Her2. There is no complete structure of Her2 and all structures currently deposited in the protein database are either of only the ectodomains or the intracellular membranes. The missing parts of the extracellular structure that connects to the transmembrane domain, is a part of/connected to the C-terminal of domain IV. This means the tail end of the structure (end of domain IV) cannot be tested for binding, as it is not the true end but just the point of cut off from the transmembrane domain/rest of domain IV. In addition the pdb structure (3WLW) contains two molecules of Immunoglobulin G (IgG) fabs that are bound to the Her2 dimer (one on each Her2 protein in the dimer). The bound IgG fabs cover some areas of the 3WLW structure, making it impossible to dock anything to the covered areas. Manipulating the structure and removing the IgG fab manually is not something that will be done in fear of that it will somehow affect Her2s structure.

3WLW is a structure of the Her2 extracellular domain homodimer. This means that it contains two Her2 (ectodomains only) proteins, that are bound to each other. It can be discussed whether or not this structure of the homodimer is correct as it is not similar to the dimer made by Her1. Nevertheless it does contain two molecules that are bound, and that will block certain areas from being searched due to steric blocking and binding. But due to the dimer conformation (head to tail), one of the two proteins have the complete domain II unhindered, and the other Her2 protein have domains I and III unhindered. While domain IV in both chains is not close to the other chain and is therefore available for docking in both of the two Her2 proteins in 3WLW. Therefore 3WLW being a dimer consisting of two Her2 proteins should not be a problem for docking purposes.

As mentioned there seems to be a major difference between the structures of the dimer presented in the 3WLW pdb as opposed to the conformation that is stated in other literature and also different than the structure proposed by Arkhipov et al. [11] as well as the structure that shows the Her1 homodimer [16] which should be similar to the Her2 one. 3WLW shows a head to tail conformation between the two Her2 proteins, with domain II being the head, and the grove between domain I and III the tail. Other literature believes the conformation of the dimer to be head to head, with domain II (specifically the dimerization arm) of both proteins interacting with each other.

In addition to the structure of the extracellular domains of Her2 given by 3WLW there are two other Her2 structures as well, 2A91[24] and 1N8Y[28]. Both of these structures have their problems and restrictions concerning docking, like 3WLW has. 2A91 is a structure of the monomer but is incomplete and only contains a few residues of domain IV, as can be seen in figure 10 (page 29).1N8Y is also a monomer of the extracellular domains of Her2, but it is not a human Her2 protein but rather Neu (Her2 from rat). Rat Her2, is an orthologue of the human Her2 and is often used in research.

In order to try to compare the structures, the chain A of 3WLW was aligned to the structures 2A91 and 1N8Y using the alignment feature in PyMOL.

Between different structures of the same protein or between close paralogs (duplicated genes in the same species) there is no need for heavy tools like DALI to make a good structural alignment. DALI is a structural alignment program that is made to make good structural alignments between proteins that does not have a large sequence similarity. When doing structural alignments for e.g. two different Her2 structures (of the same species) the sequence is 100% identical and only the positions are somewhat different. For the same protein in different species (orthologue proteins), the structure may be quite similar in some cases and others not. Tools with simpler structural alignment methods such as the inbuilt alignment feature in PyMOL will more than suffice for this purpose or for aligning with closely related structures such as rat Her2, because of the very similar structures.

After aligning 3WLW to 2A91 and 1N8Y (see results), it was decided to use 3WLW (see discussion for reasoning behind the decision). 3WLW is the preferred ligand for use in docking experiments as it's the most complete human Her2 structure found. If 3WLW can not be used, another suitable structure will be used, regardless of its species of origin.

Concerning structures of the two ligands EGF and TGF-α, not many were available. The pdb structures 1EGF [33] and 2TGF [40] were used for the two ligands. 1EGF is a structured determined by NMR and therefore contains several different conformations (in this case 16), only the first conformation of 1EGF were used for docking experiments.

Choosing search areas

General considerations and the choice of first search area

The search areas chosen for the docking experiments have many requirements for getting optimal results. Generally the goal is to have search areas that cover all the possible binding sites, and each individual search area covering all the important residues in the site while still being small enough that the run time for the docking experiments not being longer than necessary. Yet it is still important to not make them so small that no results are found. If the search areas chosen do not find any good receptor-ligand dockings more calculations will have to be run, which is time and resource consuming.

Another factor to be considered is where the molecules of interest are thought to bind, i.e. where the binding site is. A search area should be made around the binding site. It is important to note that the search area must also be designed so that the configuration that is thought to happen, have space enough to be exist within the search area. In order to theorize if there could be a binding, the residues thought to be involved in binding could be investigated. Then compare the residues and topology of the known binding ligand to the residues and topology of the ligand one wishes to dock.

Another method to investigate docking is to use a more pragmatic or statistical approach when choosing search area. This could be done by selecting the area which most often binds ligands in all proteins (or at least in all enzymes and receptor proteins). For enzymes it has been found that 83,6 % of enzymes binds their ligand(s) in the largest cleft[41]. This is also the case for receptor proteins where most of the ligands bind to the largest cleft [42]. This is because of the possibilities of interactions between residues in the receptor or ligand is the greatest when the ligand is inside a cleft in the receptor. Considering the fact that the strongest bindings would have most interactions they tend to be found in a cleft where there would be the greatest potential of numerous interactions. Therefore making a search area that covers the largest cleft of a protein will in most cases give you the best possible docking.

The first search area (gridbox) chosen was a small part of a loop that poked out of the main structures, which is the dimerization arm described in Hu et al.[12]. As it is an important area

for the dimerization process it would be logical to search this area for binding. Binding in this region could block the receptor dimerization or help it (even though current literature says Her2 homo dimerization is ligand independent). The gridbox (gridbox 1) made was rather small as the first experiment also served as a tryout to see how long time the docking experiments would take. The receptor used for the docking experiment was 3WLWa and ligand was 2TGF.

Binding of EGF and TGF- α to Her1

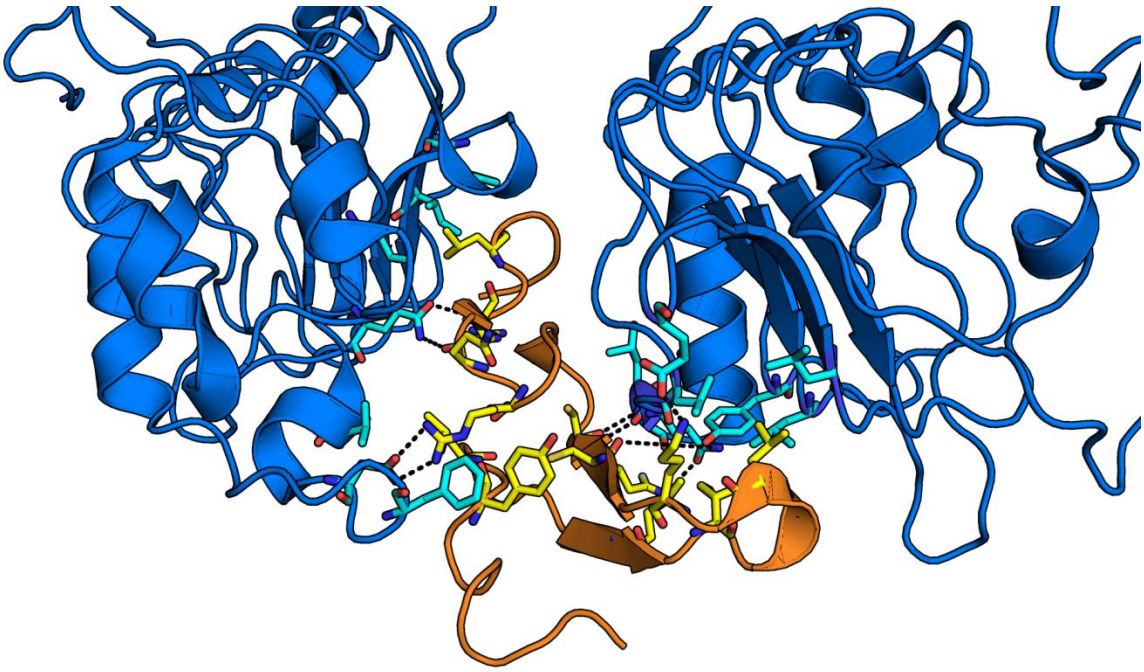

Figure 6: Showing the binding between Her1 and EGF. Domain 3 of Her1 on the left in blue, EGF in the middle colored orange and yellow and on the right is domain I of EGF in blue. Hydrogen bonds and salt bridges between the proteins are shown, with the amino acids involved depicted as sticks. Structure used for creating image is 1IVO. Image created in PyMOL [18]

The next set of gridboxes was selected based on a theory that the binding of TGF-α by Her2 resembled the binding of EGF by EGFR (Her1) [43, 44].The binding of Her2 and TGF-α could be similar to the binding between Her1 and EGF. This binding can be seen in figure 6, made with the Her1 structure 1IVO [16] which shows Her1 with EGF bound. Figure 6 shows the binding between EGF and domain I and III of Her1 where the binding happens between three binding sites (one in domain I and two in domain III)[16]. There also exists a structure for the binding of TGF-α to Her1 (pdb 1MOX [17]), as well as information about residues involved in the binding[17]. The goal is to try and make a good model for aligning and matching Her1 and Her2, and then make a gridbox for Her2 based on how Her1 binds the ligands.

Before using any methods to compare Her1 and Her2 and the binding sites, a search area was made for an area that seemed to be involved in binding by looking at the figures from Ogiso

et al.[16]. The search area contained an alfa-helix and some loops surrounding it, the search area was gridbox 2. Gridbox 2 is a search area that covers an α-helix in domain I, 2TGF was used as ligand and 3WLWa as receptor.

There is no easy method to find the same "pattern/topology", in two different structures. This causes problems for finding similarity between the binding site in Her1 and the potential binding site in Her2. Methods that compare the sequence such as sequence alignment would have difficulties finding a good alignment for the binding sites as it does not reflect the structure. While it may be possible to try to find the same residues in the same relative positions, the sequence itself cannot tell if the residues structural relation and spatial relation is the same. In addition it will be difficult to get an alignment that can be good for all 3 binding sites, as the best overall alignment may not align all the binding sites in the best way. In order to better find the binding sites in Her2, there was done a DALI alignment of a Her1 structure (1IVOa) and Her2 (3WLWa) and a manual comparison to supplement the information from the alignment. The manual comparison was done by looking at the topology that surrounded the binding residues of Her1 and trying to find the same topology in Her2, in essence comparing secondary structure and the spatial orientation of the secondary structures.

Similar topology was found and there was subsequently created suiting gridboxes for the assumed binding sites of Her2, with the data gathered from the sequence alignment of Her1 and Her2 taken into account for the creation of gridboxes. These gridboxes are gridbox 4, 5 and 6. With gridbox 4 being created for the binding site found in domain I (binding site 1 in Ogiso et al.[16]) and gridbox 5 and 6 representing the binding sites 2 and 3 respectively, which are the binding sites in domain III. The AutoDock Vina experiments will be done with these gridboxes as search area, TGF-α and Her2 structure 3WLWa. If there was found a binding in more than one of the search areas, there would be done another docking experiment to see if it can bind both, or all three at the same time, and to further explore the structure of the TGF-α and Her2 complex.

Regardless of sequence alignment being a good tool to find the binding sites in the different molecules or not, it could be good for comparing the sites between the molecules. By using the sequence alignment there could be found similarities between the binding sites in Her1 and the corresponding areas in Her2. Therefore, despite its limitations, the sequence alignments made when doing the DALI alignment was analyzed in order to find the binding sites in Her2.

Could drugs used against Her2 type cancer give a clue about possible docking locations?

There are several drugs that binds to Her2 and inhibits its involvement in breast cancer [45]. As their mechanism of action involves binding to Her2, it could be of interest to investigate wether or not any of these drugs could be working as a blocker for a ligand. There are several ways a molecule can act as an blocker on a protein, if it is an enzyme it could bind in the active site (competetive binding),or it could bind in another site and change the conformation of the enzyme (allosteric regulation). In receptors it could bind to the ligand binding site and block ligands from binding, or in Her2s case it could bind to the monomer and make a steric hindrance for the formation of a dimer. This mode of action is how the Her2 inhibitors Trastuzumab (brand name Herceptin) and Pertuzumab (brand name Perjeta) are believed to work. Nevertheless it could be of use to examine their mechanism of inhibiting and assess wether or not it could actually be inhibiting a receptor seat for EGF or/and TGF-α instead of or in addition to being a steric hindrance for dimerization.

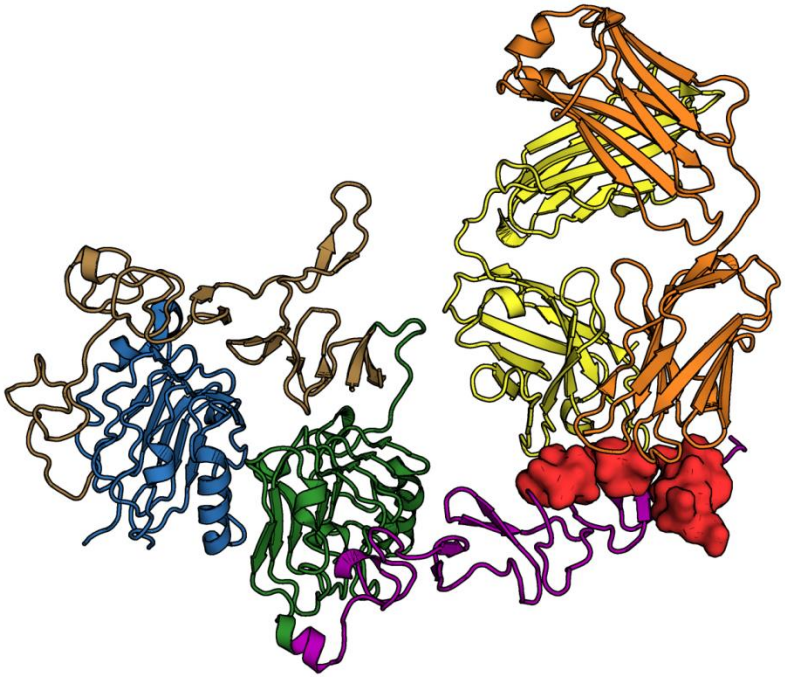

Figure 7: pdb structure 1N8Z shows Her2 ectodomain(domain I blue, domain II brown, domain III green and domain IV purple) in complex with Herceptin fab (coloured orange/yellow). Her2 residues involved in binding of Herceptin is shown with surface area colored red. The binding happens in domain IV of Her2. Image created in PyMOL [18]

While the two different Her2 blockers Herceptin/Trastuzumab and Pertuzumab/Perjeta, both bind to Her2 and blocks dimerization, they bind in different areas of the protein. Perjeta binds to domain 2 of Her2, more specifically it binds to the dimerization arm[29]. Its mode of action is blocking Her2 from dimerizing, by sterically blocking access to the dimerization arm, and

thereby blocking any potential bonds that is nescessary for a dimerization. Perjeta is a drug that works by sterically blocking Her2 homodimerization.

Herceptin/Trastuzumab, which is an immunoprotein fab, binds towards the C-terminal end of domain IV of Her2, as shown in figure 7, and is believed to cause a steric hindrance of dimerization by this binding. The mechanisms of it's inhibition of Her2 is believed to be an amalgamation of multiple reasons, including the steric hindrance of dimerization, although it is not explicitly proven to do so. But other mechanism such as hindering aggregation of Her family proteins (especially the overexpressed Her2) on the cell surface, and increasing the rate of endocytosis and degradation of Her2 (and thereby reducing available Her2 for dimerization) are believed to be involved [28]. The binding may also hinder some interactions and thereby inhibiting activation of the kinase domains. Finally, Herceptin has also been proved to hinder the truncation of the Her2 ectodomain by metalloproteases [46]. In cancerous cells, metalloproteases may be overexpressed/not inhibited as usual and be released into the extracellular space where they can cleave off the ectodomain of Her2. When this happens the kinase domain will be activated and will continously have kinase activity until its degradation. This has a large impact on the signalling systems that Her2 is involved with,as it will have a continous signal until death and thereby increasing activity of Her2 signalling. Blocking of this truncation is another reason for Herceptin being a drug that hinders tumor growth for certain cancers. Overall the binding of Herceptin induces many changes that decreases the impact of overexpression of Her2, and its binding is interesting with concern to the binding of EGF and TGF-α.

Therefore a search area was made in order to check this, but unfortunately the domain IV of the 3WLW structure is cut short of the residues that bind Herceptin. The structure made by Cho et al. [28] to show the binding of Herceptin to a Her2 monomer (pdb 1N8Z) could also be a possibility to use as receptor in docking experiments. But can not be used because of the prescense of Herceptin in the structure would disturb docking attempts. Therefore the structure of rat Her2, pdb 1N8Y, was used as it contains all the needed residues in domain IV. The search area covers all the residues that is involved in the binding of Herceptin, shown in figure 7 with their surface area shown and coloured red.

Gridbox 6 was designated as the search area for the area of binding between domain IV and Herceptin, and gridbox 6 was run in AutoDock Vina with 1N8Y as receptor and TGF-α (2TGF) as ligand.

IgG fab binds to Her2

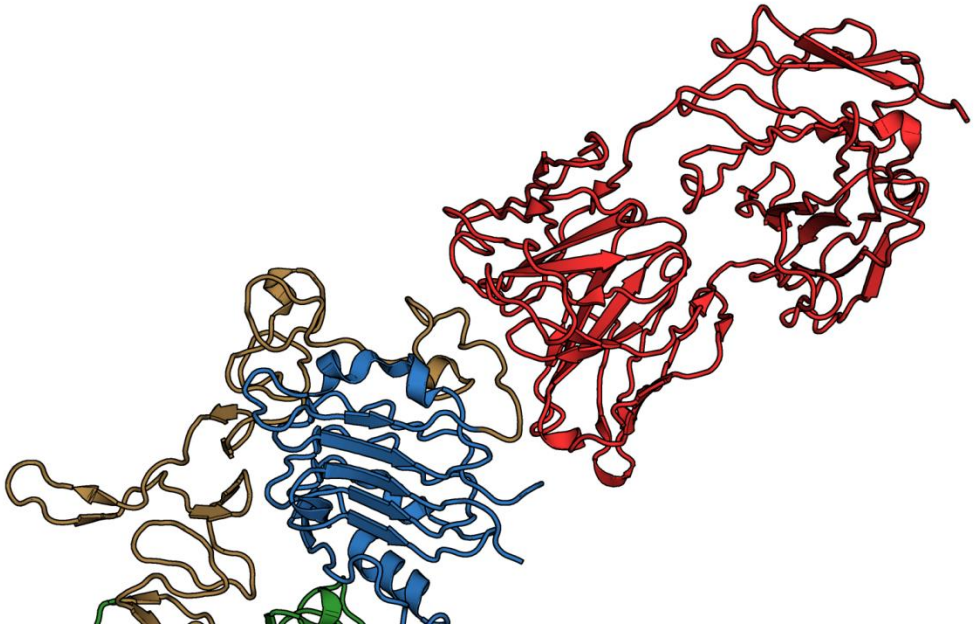

Figure 8: IgG fab (red) shown bound to a loop in domain II of Her2 (brown). Structure used is 3WLWa. Domain I (blue) may also be involved in the binding of the IgG fab. Image created in PyMOL [18]

Before the protein docking experiments of this thesis started, a trial run of docking TGF-α (2TGF) on Her2 (3WLW) was done by professor Lars Skjeldal at the Norwegian University of Life Sciences[47]. TGF-α was successfully docked to the IgG fab, on the end part of the IgG fab opposite of its binding to Her2.

As mentioned earlier one of the downsides of the pdb structure 3WLW was that the IgG fab bound to Her2 restricts docking in the area of the IgG binding, and also the are surrounding the binding. But if the IgG fab can have a stable binding to Her2, maybe it is possible for TGF-α to bind there as well. The specifics of the Her2-IgG fab binding is not discussed in the article of the 3WLW structure [12] and there is therefore no information to theorize if the type of bindings found could be possible in TGF-α, but this is not a big enough reason to discourage from trying. As can be seen IgG is bound to a loop area in domain II of Her2, pointing away from the rest of the protein, this loop in domain II is therefore interesting when it comes to bindings to Her2.

A search area was therefore created that included the loop in domain II as well as space to allow the ligands to bind in a similar manner to IgG fabs binding to Her2. The search area that was made, was given the number 7, making it gridbox 7. The structure used to do this search will be 1N8Y. 3WLW could not be used as it has a IgG fab bound in the search area, therefore 1N8Y was decided to be used as it has no steric blocking and is well suited for docking as it had been used in previous docking experiments. Gridbox 7 is a search area that

covers a loop in the beginning of domain II of Her2, the ligands for the experiment will be pdb structures 2TGF and 1EGF and receptor will be 1N8Y.

Adjusting the search areas

From the results of the Vina docking experiments shown in the figures 12 and 13 on page 31 and 32, it is visible that the docking software has modified the TGF-α structure in order to make it fit inside the search area. In addition the ligands structure overlap with the structure of the receptor, in some places having two atoms in almost the same position. This is because the search area did not include space for the entirety of the TGF-α structure to dock to Her2. The search area has to be expanded in order for the ligand (TGF-α) to fit inside it while also including the residues of Her2 that is believed to be involved in the binding. Therefore the search areas/gridboxes for vina experiment 3-7 will be remade, and new docking experiments will be done with the revamped gridboxes.

As there were no results for gridboxes 1, 3, 4 and 5 and poor results from the gridboxes 2 and 6, something may have been wrong with the gridboxes, or their placement has been poor. After much deliberation, the fault was determined to be the size of the gridboxes. They were too small to accommodate the ligands inside themselves. New gridboxes were made for the gridbox 6 with space for the ligand, as well as the yet to be run gridbox 7 was adjusted to accommodate this as well. The gridbox 6 that was first run and was too small, was defined as version 1 (v1), and the new, more spacious one, was defined as version 2 (v2). The gridboxes 3, 4 and 5 were made into one new gridbox named gridbox 8. Gridbox 1 was remade into gridbox 9, this time larger and also was made to include the possibilities of a binding similar to the binding of Perjeta to Her2.The 2nd gridbox was not remade as it was in the same area as the gridbox 3 which was now accommodated to a new one (gridbox 8), and would be superfluous. Overall all areas of interest in Her2 that has been searched up to this point, will be searched again with new, larger gridboxes.

The gridboxes 3, 4, and 5 was made to mimic the binding sites from Her1[16] in Her2. They were too small to fit the ligands into them. Making new gridboxes for each of the sites, which would also accommodate the ligands, would result in significantly larger gridboxes. This especially as the original search areas was made with the intent that they should investigate if there could be a good binding of the ligands, with the ligand in an angle that wouldn't be similar to the one that was found in pdb structures 1IVO and 1MOX and shown in figure 6. Gridboxes that would account for these angles would need to be very large compared to one that just assumed the binding would be similar to the known binding of Her1 and EGF/TGF-α. Therefore it was decided that it would be better to make one gridbox that covers the binding sites, and the area corresponding to the binding site between Her1 and the ligands. In addition this new search area will take into account the information of the binding between TGF-α and Her1 which the gridboxes 3, 4, and 5 did not. The new search area will therefore be more suited to dock TGF-α than the ones created only from knowledge about the Her1:EGF interactions.

The choice of receptors for these new gridboxes was the same as the originals that they are an improvement of. Gridbox 6 uses 1N8Y as it needs a domain IV that is as complete as possible, and 3WLW has a much shorter domain IV than 1N8Y. The choice of the two different Her2 chains for gridbox 8 and 9 is due to the head to tail conformation of the Her2 dimer in 3WLW. The dimerization arm of 3WLWa is located in the gap between domain I and III for 3WLWb. Gridbox 8 can therefore not use 3WLWb as there is steric hindrances in the gap between domain I and III, while gridbox 9 can not use 3WLWa as the dimerization arm is sterically blocked.

In summation, gridbox 6 was recreated with a larger search area but covering the same residues. The ligands for this search area were 1EGF and 2TGF with receptor 1N8Y. Gridbox 3,4 and 5 were combined to a new search area, gridbox 8, using 1EGF and 2TGF as ligands with 3WLWa as receptor. A new search area was created around the dimerization arm (gridbox 1 also covered this area), using a larger search area than previously. This new search area was gridbox 9 and using 1EGF and 2TGF as ligands, with 3WLWb as receptor.

Summation of the search areas used in AutoDock Vina experiments

In total 10 different gridboxes were used in docking experiments were used in this thesis. They were all different, and even though some were in the same area, they differed in size and/or coordinates. The area the different gridboxes cover are shown in figure 9, and the gridboxes are noted next to the area they are found. The area is representative of where the gridboxes are but does not represent the gridboxes size or the excact coordinates.
In addition figure 9 is a two dimensional image made to represent searcha areas that exists in three dimensions and therefore does not show the total extent of the search areas. The coordinates and size of the gridboxes are given in Appendix 2, and can be used together with the program AutoDock Tools to recreate the search areas and view them in a interface that imitates three dimensions. In figure 9 an overview of all the search areas can be seen.

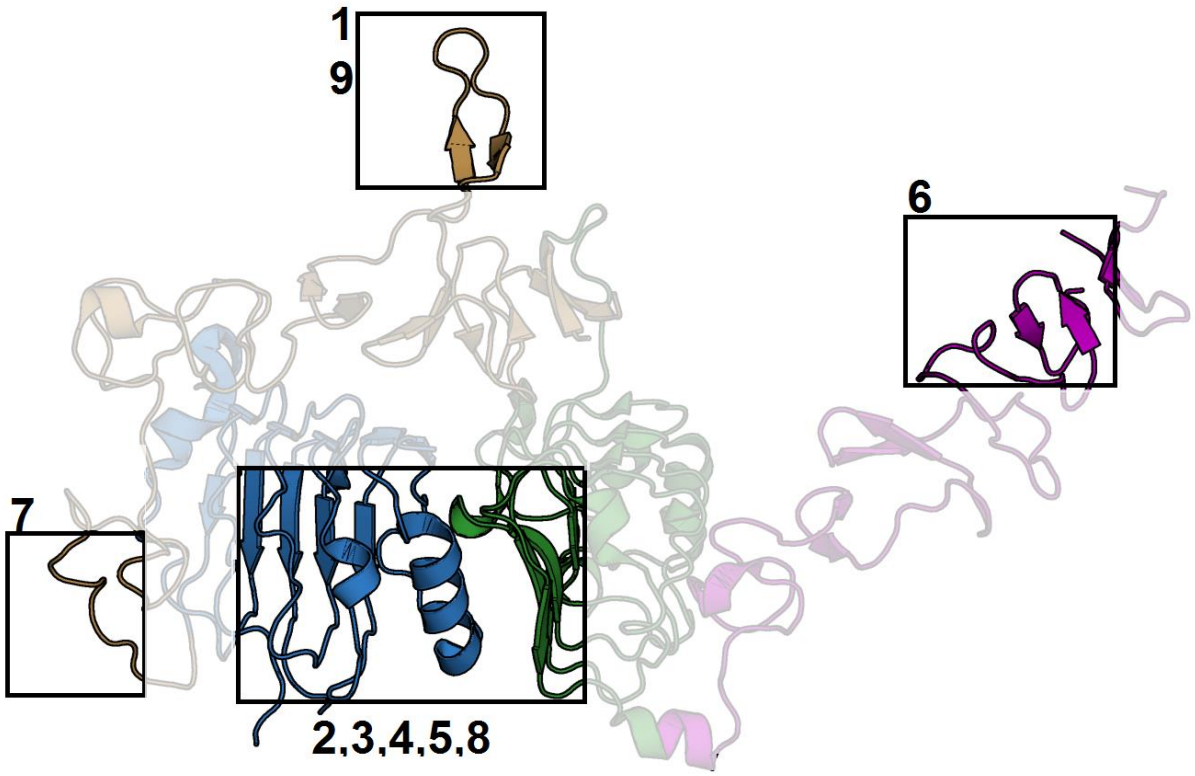

Figure 9: Overview of the different gridboxes. The numbers represents which gridboxes can be found inside the boxed off areas. The area is just a representation and is not identical with the real gridboxes. Note that both gridbox 6 v1 and 6 v2 is in the same area denoted by the number 6, as the difference is just size. Original image created in PyMOL [18].

Assesment of the quality and validity of results from docking experiments

In order to be assured that the results obtained from the docking experiments are of good quality, the data and resulting structures has to be interpreteted. Vina gives two main results that has to be interpreted, the binding affinity and the docked structure. The affinity can be used as a measurement of how good the docking is and does not need any kind of processing, opposed to the docked structures that do need interpretation. In the interpretation the goal would be to asess the docking, and find where the bindings and interactions between the two proteins are. Therefore the docked structures are inspected and interpreted.

Theoretical bindings

The general area where the docked structures are close to each other (and thus where majority of bonds/interactions will take place) was found by inspecting the results (structures between ligand and receptor) from AutoDock Vina. In order to try and get more information out of the docking and to try and create a model of the binding that includes interactions and bonds, potential bonds and interactions will be theorized. These bonds are assumptions on what bonds exists in a binding between the ligand and receptor based on structures given by the docking program, and are not to be taken as definitive, as they are just assumptions.

The theorized bonds between the docked ligands and receptors are done by observing the structure in a structure visualization program, in this case PyMOL. Observing the structure the residues of the ligand and receptor that are close to each other are observed and if there is possibility of a type of bond or interaction. Note that any form of covalent bond is not considered as AutoDock Vina does not consider them, and therefore they will not be used to create a model of the bonds as that would not be accurate. The distance between the atoms/groups in question is measured, and if the distance is too large (in this case over approx. 5Å) the bond is not seen as viable and therefore not included. All the bonds or interactions are denoted and presented in the results. As this is a method that relies on observations by a person it is biased and prone to human error that could lead to bonds and interactions going unnoticed. Especially for hydrophobic interactions as they can be difficult to spot, because they do not have two easy to spot suspects (a close hydroxyl/carbonyl/carboxylic acid group and an amino group). The measurement is kept in PyMOL and used to depict a bond when making figures.

Ramachandran plots of AutoDock Vina results

In addition to interpreting the structures and its interactions and bonds one other test on the ligand structures will be done in order to further be assured that the results are authentic. In order to find out if the structure of the ligands was changed by AutoDock Vinas flexility options, to a conformation that would be unfavourable. A Ramachandran plot was done for

each of the docked structure as well as the input structure of the ligands. This was done using a webtool for creating several different informational figures and plots by Duke University School of Medicine's Biochemistry department[39], the tool is available at http://molprobity.biochem.duke.edu/ .When making the Ramachandran plots the output ligand structures were converterd from a pdbqt file to a pdb file only containing one conformation of the ligand (the one with highest binding affinity), this was done with AutoDock Tools. After uploadeding the structure, hydrogens were added to the structures by using the inbuilt tool on the webtool. The recommended options given by the tool were always used, changing no options from the default. After adding hydrogens, Ramachandran plots were made by the web tool, for each of the structures gained from the docking experiments.

"Test docking"

Another procedure that will be performed to assure the authenticity of the results, is docking TGF-α to other proteins than Her2. Thereby, possible program biases in docking of large proteins (such as Her2) can be checked, or check if TGF-α is a protein that binds easily. Unfortunately in the event of a positive docking result there is no way of seperating which of these two issues causes TGF-α (and most likely EGF) to be so easily docked with Vina.

In order to do this two α/β protein structures consisting of over 400 residues were selected randomly from the protein database. There were then selected gridboxes that encompassed random parts of the surface with the search area being 36×36×36 Å.This size was chosen, because the volume of this are is similar to the volume of the search areas used in the successful docking experiments. The two protein structures which was selected were 1WP6 and 2VJL. 1WP6 is a structure of an amylase, which is a type of enzyme that breaks down polysaccharids into smaller sugars.In this case the amylase converts polysaccharids to maltohexaose [48]. 2VJL is also a structure of an enzyme, namely a transferase that transfers CoA between specific molecules[49]. Due to time constraints only two search areas for one of the proteins (1WP6) were used for docking experiments, the search areas used are listed in Appendix 2.

Results:

Structural alignments

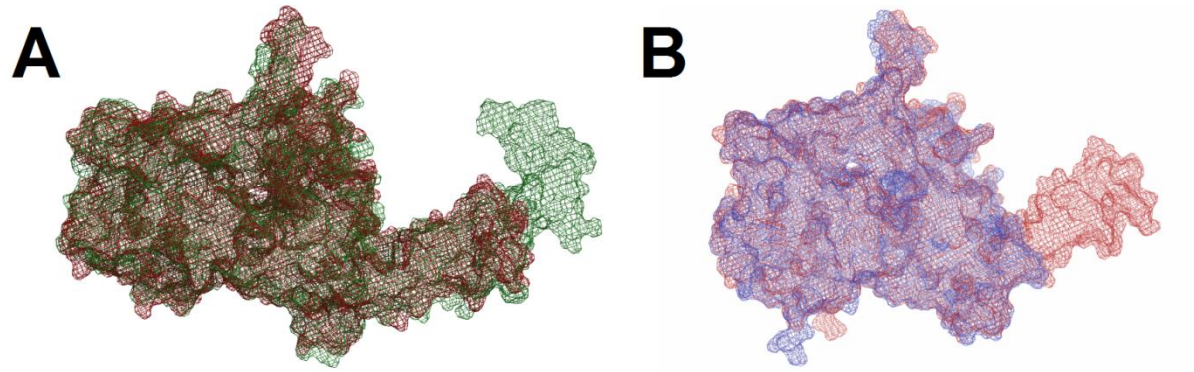

Figure 10: A) 3WLW (red) aligned to 1N8Y (green). **B)** 3WLW (red) aligned to 2A91(blue) Shown in mesh mode from PyMOL, which represents the surface. Image created in PyMOL [18]

An alignment was done of the structures of 1N8Y and 2A91 to the structure of 3WLWa. This was done in order to see where the differences were between the models. Only chain A of the dimer structure 3WLW was included in the alignment to the complete structures of 1N8Y and 2A91. The alignment was done with PyMOLs alignment feature. As can be seen in figure 10 there are some easily seen differences between the two structure alignments. It can be seen that the rat Her2, 1N8Y, has a larger domain IV, and that 2A91 does not have the majority of domain IV included in the structure. The difference in domain IV most likely reflects different cut off sites from the transmembrane domain, and not difference in size of domain IV between human and rat Her2. Even though the difference may not be observed in the true structure of the whole Her2 protein, it's still important to consider these differences in the structures. There were no other major differences found by inspecting the surfaces of the proteins.

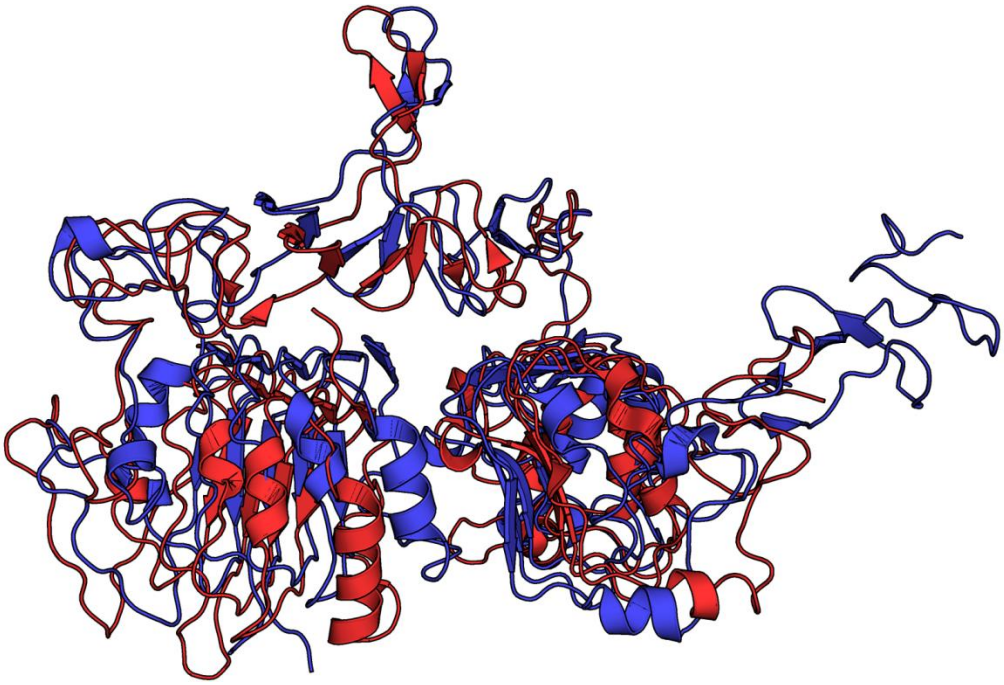

Figure 11. Alignment of Her1(1IVOa) in red and Her2 (3WLWa) in blue, using DALI[36] . Image created in PyMOL [18]

A structural comparison using the DALI tool on the structure 3WLWa (Her2) and the structure 1IVOa (Her1, in an active conformation) was performed. The DALI server provided several alignments (see appendix 1), the one with the lowest Z-score was chosen. The alignment with highest Z-score is depicted in figure 11.The structural alignment shows that the structure of Her1 and Her2 is very close, with the type of secondary structures being almost identical. The largest difference is that the positioning (in the room) of the two proteins is not identical. Along with the DALI structural alignment, a sequence alignment that takes secondary protein structure into accord was provided and can be found in appendix 1.

The Z-score for the alignment is 33.3 and the alignment's RMSD is 5.The Z-score is high which indicates that the sequence alignment is good whereas the RMSD is not as low as one wants. The value is not high enough to think there is something wrong with the alignment and coupled with the high Z-score, it indicates a good alignment. In this alignment it is possible to compare residues to each other both with regards to which amino acids it is, and the amino acids chemical properties as well as which type of secondary structure it is in. Together these two alignments provided by the DALI tool can be used to try and replicate the bindings and contacts that was found between TGF-α /EGF and Her1.

Results from Vina

Docking results from gridbox 1:

AutoDock Vina returned no docked structure.

Docking results from gridbox 2:

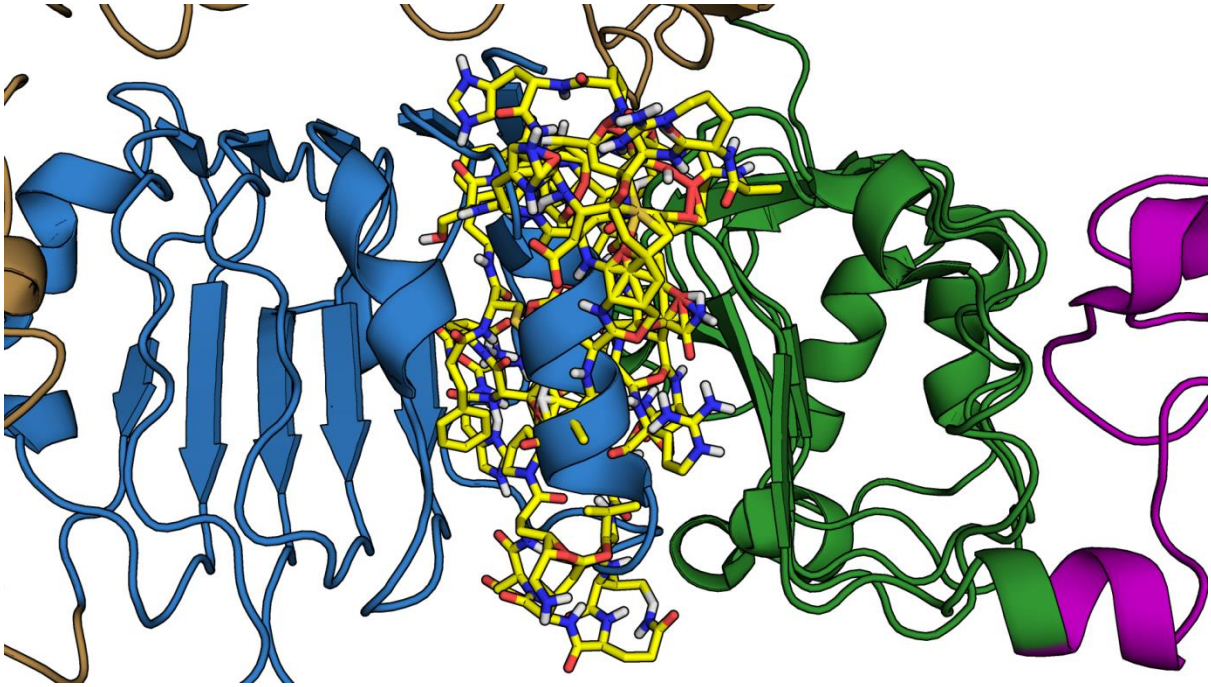

Figure 12: TGF-α (yellow) docked to domain I (blue) of Her2 structure 3WLWa. It can be clearly seen that the structures of TGF-α and Her2 is overlapping. Domain II is in brown, domain III green and domain IV in purple. Image created in PyMOL[18]

There was found a binding for the second run of Vina (Gridbox 2). The binding affinity of the two structures with TGF-α docked to Her2, is given in table 1 found in Appendix 3. Because of the positive values of the affinity, there can not be a binding, and the findings done by Vina are in this case, a false positive. In addition when looking at the structure output(depicted in figure 12), the TGF-α is warped and it's also docked on to the Her2 molecule in such a manner that most of the atoms overlap. The findings from this run of Vina are impossible in regards to both physical chemistry and stereochemistry.

Docking results from gridboxes 3, 4 and 5

AutoDock Vina did not find any docked structures for the gridboxes 3, 4, and 5 using TGF-α as the ligand to the receptor 3WLW.

Docking results from gridbox 6, v1

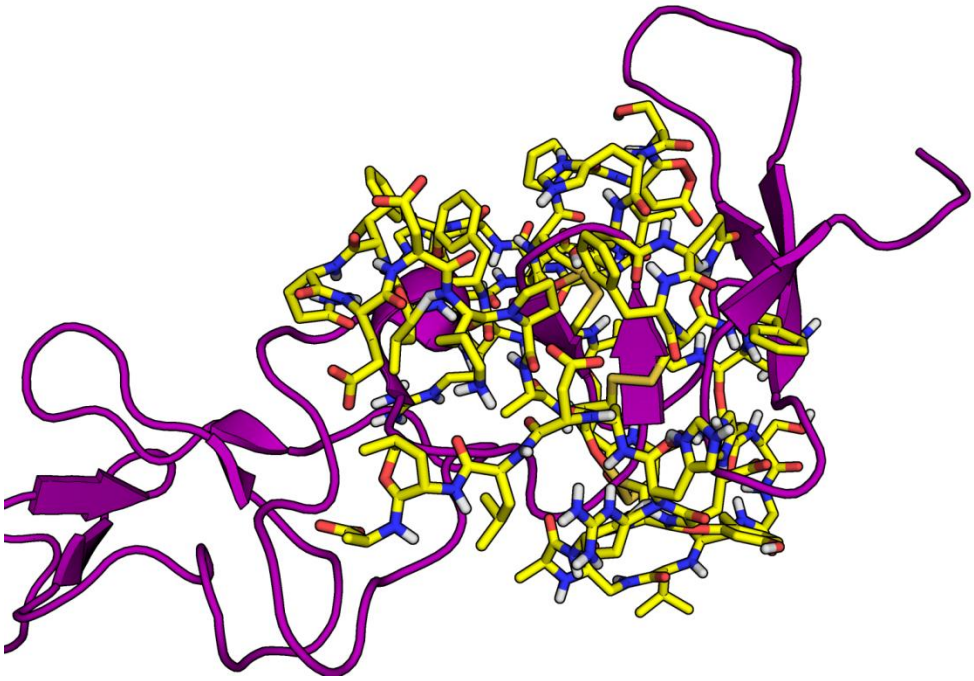

Figure 13: TGF-α (yellow) docked towards the end of domain IV (purple) of Her2 structure 3WLWa. It can be clearly seen that the structures of TGF-α and Her2 is overlapping. Image created in PyMOL[18]

There was found one weak binding of TGF-α to Her2 structure 1N8Y for gridbox 6 v1. The positive value of the binding affinity and the fact that the structure of the docked ligand overlaps with the structure of the receptor indicates that this is a false positive from the docking experiment.

Docking results from gridbox 6 v2

The docking experiments with gridbox 6 v2 gave 9 results for both ligands. All with negative binding affinity, as shown in table 3 and 4 in appendix 3.

From the docked structure of the top result from TGF-α, it is easily seen that this can be a good binding with many bindings and interactions involved. There are three main sites for binding similar to what was found between Her2 and Herceptin. These three sites are residues 559-562 for the first one, 571-573 for the second and for the last are residues 582-592. Residues 41,42,45 and 46 of TGF-α are candidates for binding to the first site in Her2. For the second site residues 9-11 in TGF-α are of interest, and for the last site the residues close to the N-terminal, specifically the two first residues, are close to the later residues of site 3, and residues 25-28 in TGF-α are close the beginning of site 3 in Her2.

EGF is docked to Her2 in a similar manner to the docked TGF-α and also have the same 3 binding sites that the docked TGF-α had. Those three sites being residues 559-562 of Her2 making up the first site, residues 571-573 and 604 for the second and 582-592 for the third. From EGFs side residues 4-6 could be involved in site 1 and residues 18-20 for site 2. Site 3 is split up just as for the docking of TGF-α, residue 16 and 17 of EGF are close residue 592, while residues 44-46 of EGF are close to residues 585-587 of Her2.

The theorized bindings between the docked ligands to the Her2(1N8Y) for gridbox 6 v2 are given on the next page.

Concerning all theorized bonds presented in this thesis:

All theorized bonds are from the docked structures with thea highest binding affinity. All bonds are theorized and may not represent the true bonds, and some bonds may have been overlooked.

The bonds are written in this format:

Residue in Ligand, group interacting – Residue in Her2, group interacting (Type of interaction)

This is the case for all theorized bonds and interactions in this thesis.

The bonds theorized between Her2 (1N8Y) and TGF-α (2TGF)

Val1, mainchain CO group – Ile592, mainchain NH group (Hydrogen bond)
His4, sidechain amino group – Asp571, sidechain COOH group (Hydrogen bond)
His4, sidechain amino group – Asp571, mainchain CO group (Hydrogen bond)
Asp10, mainchain NH group – Ser572, sidechain hydroxyl group (Hydrogen bond)
Val25, sidechain – Pro585, sidechain (Hydrophobic interactions)
Val25, sidechain – Met590, sidechain methyl groups (hydrophobic interactions)
Gln26, sidechain NH2 group – Leu587, mainchain CO group (Hydrogen bond)
Gln26, mainchain NH group – Ser588, sidechain hydroxyl group (Hydrogen bond)
Arg42, 1st sidechain amino group – Asp561, mainchain CO group (Hydrogen bond)
Arg42, 1st sidechain amino group – Ser573, sidechain hydroxyl group (Hydrogen bond)
Arg42, 2nd /3rd sidechain amino group – Asp561, mainchain CO group (Hydrogen bond)
Arg42, 2nd /3rd sidechain amino group – Ser573, mainchain CO group (Hydrogen bond)
His45, sidechain amino group - Asp561, sidechain COOH group (Hydrogen bond)

The theorized bonds are visualized in figure 14, which is a figure that shows TGF-α docked to domain IV of Her2.

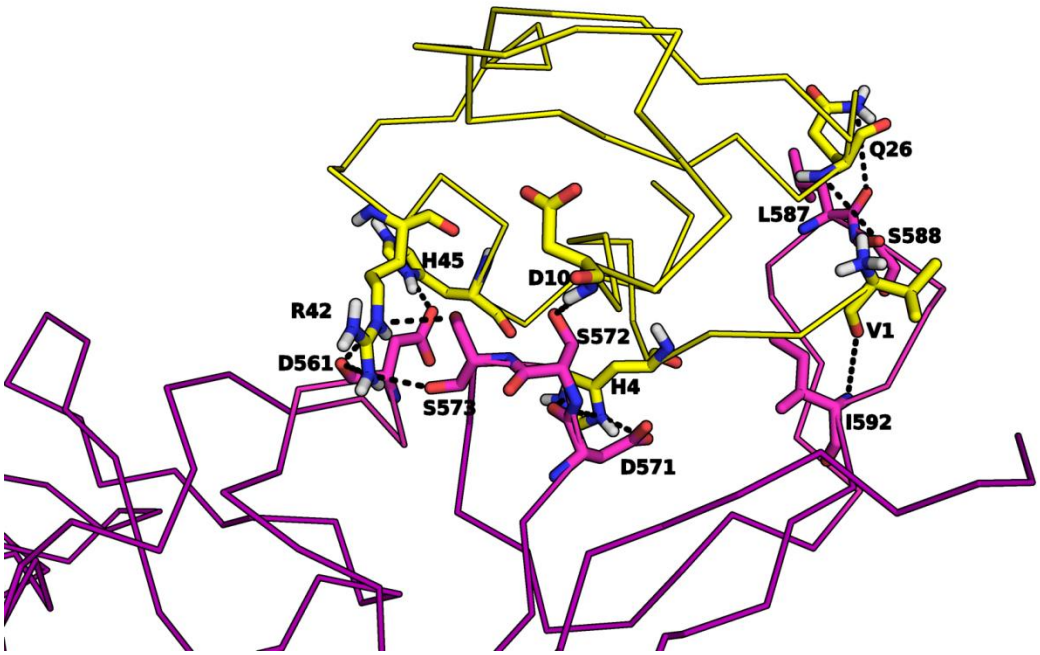

Figure 14: TGF-α (yellow) docked to domain IV of Her2 (purple), the N-terminal of Her2 can be seen furthest to the right in figure. The residues involved in bindings are shown in "sticks", with dashes between them representing different type of bonds. Hydrophobic interactions have not been included in this figure. Residues are labeled using the one letter name for amino acids. Image created in PyMOL [18]

The bonds between Her2 (1N8Y) and EGF (1EGF)

Asn1, sidechain amide group – 562Gln, sidechain amide group (salt bridge)
Asn16, sidechain NH2 group – Pro591, mainchain CO group (Hydrogen bond)
Cys20, mainchain CO group – Ser573, mainchain NH group (Hydrogen bond)
Cys20, mainchain NH group – Ser572, sidechain hydroxyl group (Hydrogen bond)
Arg45, sidechain 2nd /3rd amino group – Met590, mainchain CO group (Hydrogen bond)
Asp46, sidechain COOH group – Ser588, mainchain NH group (Hydrogen bond)

EGF docked to domain IV of Her2 can be seen in figure 15, where also the theorized bonds are depicted.

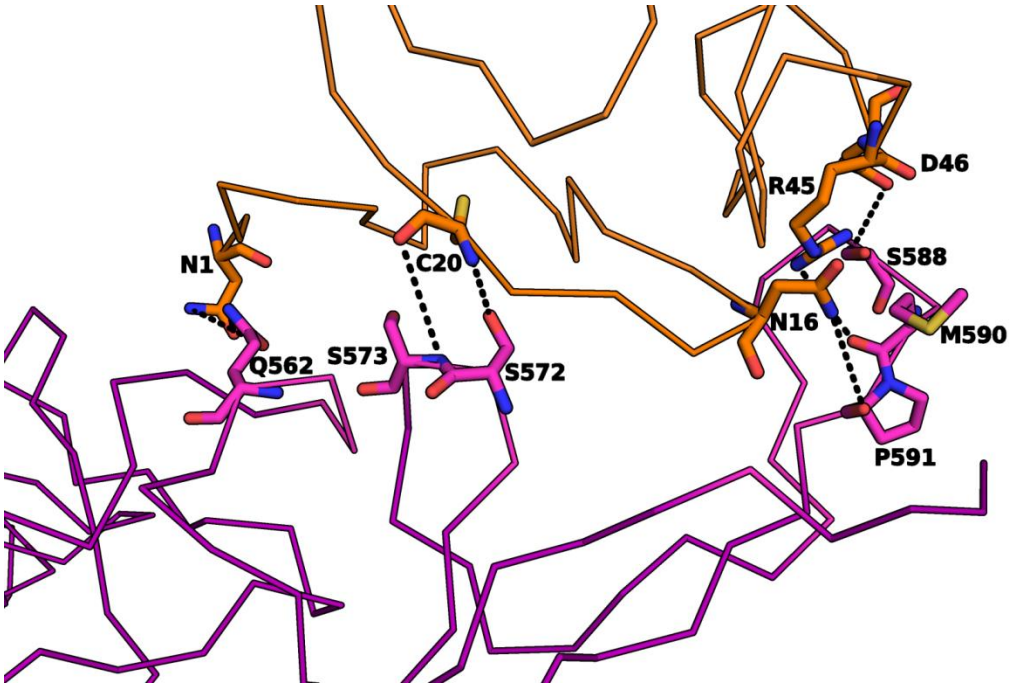

Figure 15: EGF (orange) docked to domain IV of Her2 (purple), the N-terminal of Her2 can be seen furthest to the right in figure. The residues involved in bindings are shown in "sticks", with dashes between them representing different type of interactions and bonds. Residues are labeled using the one letter name for amino acids. Image created in PyMOL[18]

Docking results from gridbox 7

With search area defined as gridbox 7, several docked structures from docking experiments between TGF-α/EGF and Her2 were found. There was given 9 different dockings for both ligands, the binding affinities can be seen in tables 5 and 6 in appendix 3. All the 9 affinities were negative for both ligands.

The docked structure with highest affinity for TGF-α docking to Her2, shows TGF-α wrapped around the part of Her2 that was searched. For Her2 it is mainly the residues from the start of domain II that are near TGF-α, this being residues 167-173 as well as residues 186 and 187. The latter two are very close and does have some interaction/bond with residues 26 and 27 of TGF-α. The residues of TGF-α that are close to the aforementioned residues are residues 4, 24, 27, 35, 36 46-50. In addition the N-terminal of TGF-α is very close to residues 117 and 118 of domain I in Her2.

For the EGF docking, the ligand has docked in such a way that it only has contact with domain II specifically the long loop that consists of residues 167 to 177. These residues seem to be in contact with residues from different parts of EGF, and the ligands structure is therefore in such a way that it allows many residues to be near the residues of Her2 involved. The EGF residues that could be involved are residues 9, 10, 26-28, 48, 49, as well as residue 52 though this residue may actually be interacting with residue 186 of Her2 rather than any of the residues mentioned above.

Based on the observations and a closer inspection the following bindings were suggested:

The bonds theorized between Her2 (1N8Y) and TGF-α (2TGF)

His4, sidechain amino group – Cys171, mainchain CO group (Hydrogen bond)
Leu24, sidechain – Pro173, sidechain (Hydrophobic interaction)
Val25, sidechain – Pro172, sidechain (Hydrophobic interaction)
Gln26, mainchain NH group – Glu186, sidechain COOH group (Hydrogen bond)
Gln26, sidechain NH2 group – Ser187, sidechain hydroxyl group (Hydrogen bond)
Glu27, sidechain COOH group – Pro176, mainchain NH group (Hydrogen bond)
Ser36, mainchain hydroxyl group – Arg167, sidechain amino group (Hydrogen bond)
Ser36, sidechain carbonyl group – Arg167, sidechain amino group (Hydrogen bond)
Asp47, sidechain COOH group – Ala170, mainchain NH group (Hydrogen bond)
Leu49, mainchain CO group – Arg167, mainchain NH group (Hydrogen bond)
Ala50, mainchain COOH group* – Arg167, mainchain NH group (Hydrogen bond)
Ala50, mainchain COOH group * – Arg167, sidechain amino group (Hydrogen bond)
*Ala50 is the C-terminal residue and therefore has a COOH group in its mainchain as it is not involved in a peptide bond.

The structure of TGF-α docked to Her2 in domain I and II, is depicted in figure 16. In the figure the theorized bonds are also visualized and the residues involved in them labeled.

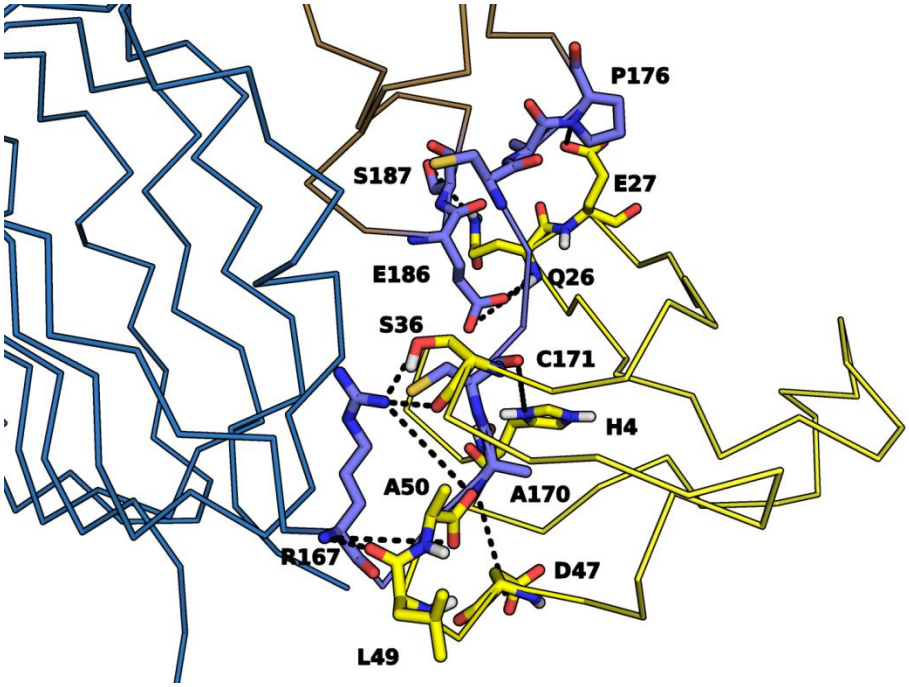

Figure 16: TGF-α (yellow) bound to domain I (blue) and II (brown) of Her2. Theorized bindings are shown with dashes between the residues, which are shown in sticks. All Her2 residues shown as sticks have been colored blue regardless if they are in domain I or II. Image created in PyMOL [18]

The bonds between Her2 (1N8Y) and EGF (1EGF)

Ser9, mainchain CO group – Ala170, mainchain NH group (Hydrogen bond)
Ser9, sidechain hydroxyl group – Arg167, sidechain amino group (Hydrogen bond)
Tyr10, sidechain hydroxyl group – Pro173 mainchain NH group (Hydrogen bond)
Asp27, mainchain CO group – Cys174, mainchain NH group (Hydrogen bond)
Arg48, mainchain CO group – Arg169 sidechain amino group (Hydrogen bond)
Arg48, mainchain NH group – Ala170 mainchain CO group (Hydrogen bond)
Arg48, sidechain amino group – Ser168 mainchain CO group (Hydrogen bond)
Trp49, sidechain amino group – Glu186, sidechain COOH group (Hydrogen bond)
Arg53, mainchain COOH group* – Glu186, sidechain COOH group (Hydrophilic interaction)
*Arg53 is the C-terminal residue and therefore has a COOH group in its mainchain as it is not involved in a peptide bond.

The structure of EGF docked to Her2 in domain I and II can be seen in figure 17. In the figure the theorized bonds are also visualized and the residues involved in them labeled.

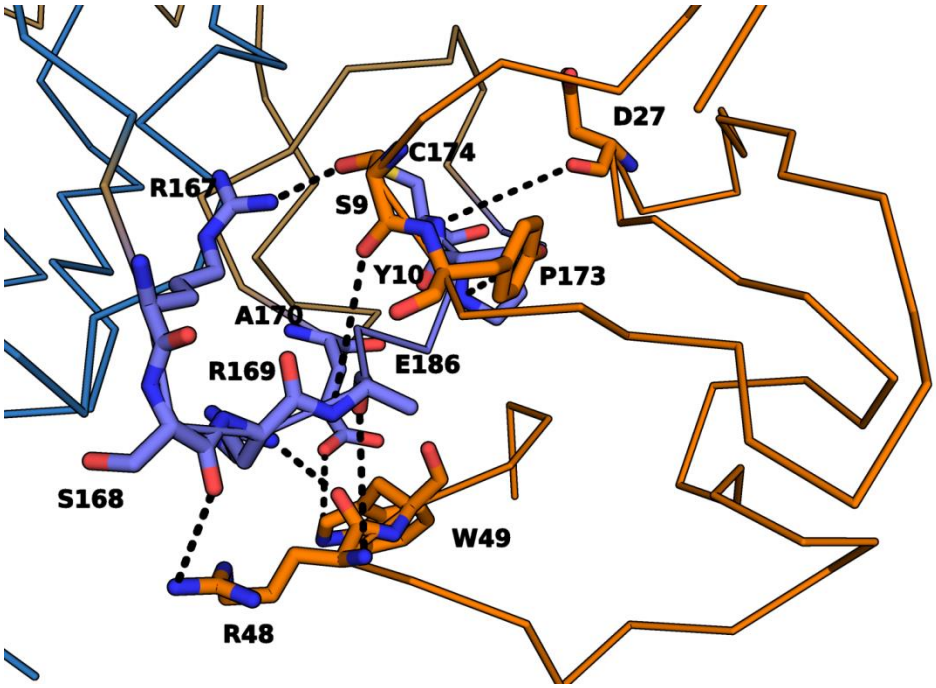

Figure 17: EGF (orange) bound to domain I (blue) and II (brown) of Her2. Theorized bindings are shown with dashes between the residues, which are shown in sticks, hydrophobic interactions not included in model. All Her2 residues shown as sticks have been colored blue regardless if they are in domain I or II. Image created in PyMOL[18]

Docking result from gridbox 8

There were found 9 possible docking conformations for both TGF-α and EGF to Her2 (3WLW) for gridbox 8. The data can be viewed in tables 7 and 8 in appendix 3.

The best resulting docking of TGF-α has TGF-α in between the domain I and III athwart to the direction of the domains of Her2 with the C terminal towards the alfa helix of domain I, and the N terminal in the middle of the pocket between domain I and III . Concerning interactions it seems that they are mainly between the residues around the C-terminal of TGF-α and the loop between the 3rd beta sheet and 2nd alfa helix of domain I in Her2 (approx. residue 41-45), as well as some interactions between the residues 4-7 of TGF-α and the residues found directly below the three last beta sheets in domain III (residues 424, 448 and 475). There also appears to be some interactions between the N-terminal of Her2 and one of the residues around residue 40, and between residues 25-29 of TGF-α and some of the residues between residues 255 and 364 of Her2 (the area close to the 2nd binding site of TGF-α/EGF to Her1).

The docked EGF structure is also placed athwart to the direction of the Her2 domains, in the pocket between domain I and III. With the C terminal almost poking out on the side of Her2 where domain IV is found and the N terminal placed towards the end of the pocket on the other side. The N terminal is in fact buried in a pocket, but not the hydrophobic pocket of domain III but rather another pocket in domain III. Because the N-terminal is buried in a pocket that is in the middle of the side of domain III that points toward the pocket between domain I and III all of the residues in EGF that is close to the N-terminal is close to domain III of Her2 and could have interactions/bonds. The residues 8-10 that are very close to residues 362-365 which correspond to one of the binding sites described in Ogiso et al. (2012). Further there seems to be similar bindings between the docked TGF-α and EGF, as EGF seems to also have a binding to the residues in between the 3rd beta sheet and 2nd alfa helix of domain I in Her2 (as stated earlier this is approx. residues 41-45 of Her2), the residues involved in this binding from EGF are residues 44-49 . There may also be some interactions/bindings between the residues 24-27 of EGF and 424 + 448 of Her2 (in the area of the central beta sheets of domain III.

Based on the observations above and a closer inspection the following bindings were suggested:

The bonds theorized between Her2 (3WLW) and TGF-α (2TGF)

Asn6, mainchain NH group – Gly425, mainchain CO group (Hydrogen bond)
Asp7, mainchain NH group – Gln424, mainchain CO group (Hydrogen bond)
Asp7, mainchain CO group – Gln424, sidechain NH2 group (Hydrogen bond)
Asp7, sidechain COOH group – His448, sidechain amino group (Hydrogen bond)*
Asp7, sidechain COOH group – His448, sidechain amino group (Hydrogen bond)*
Cys8, mainchain CO group – Gln424, sidechain NH2 group (Hydrogen bond)
His12, sidechain amino group – Glu357, sidechain COOH group (Hydrogen bond)
Asp 28, sidechain COOH group – Pro363, mainchain NH group (Hydrogen bond)
Asp 28, sidechain COOH group – Gly361, mainchain NH group (Hydrogen bond)
Asp 28, mainchain NH group – Asp360, sidechain COOH group (Hydrogen bond)
His35, sidechain amino group – Glu18 sidechain COOH group (Hydrogen bond)
Val39, sidechain – Leu101, sidechain (Hydrophobic interaction)
Arg42, 2nd/3rd sidechain amino group – Asp99, mainchain CO group (Hydrogen bond)
Arg42, 2nd/3rd sidechain amino group – Gly98, mainchain CO group (Hydrogen bond)
His45, sidechain amino group – Asp99, sidechain COOH group (Hydrogen bond)
Ala50, mainchain NH group – Pro17, mainchain CO group (Hydrogen bond)

*There are two available oxygens in a COOH group each one in this case has the possibility of a hydrogen bond to one of the two amino groups in the Histidine (one amino group for each Oxygen)

In figure 18, TGF-α is shown docked in the pocket between domains I and III of Her2, with the theorized bonds included to show the strength of the binding.

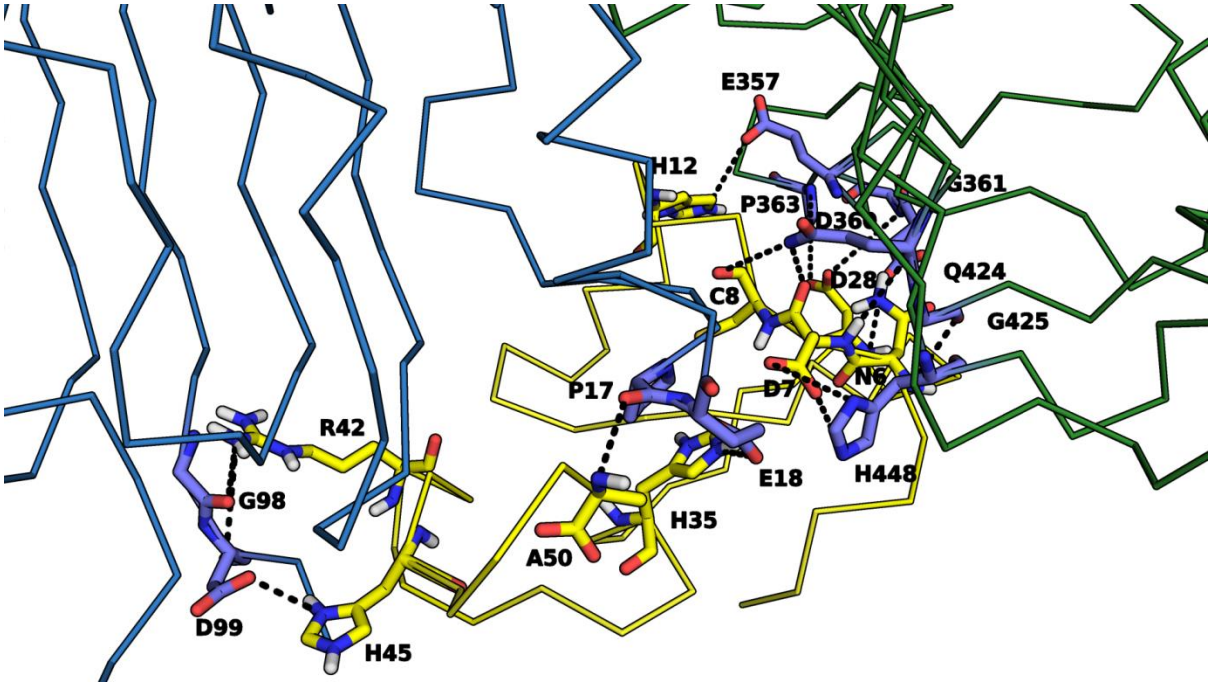

Figure 18: TGF-α (yellow) docked between domains I (blue) and III (green) of Her2. Theorized bindings are shown with dashes between the residues, which are shown in sticks. Hydrophobic and hydrophilic interactions not shown in figure. All Her2 residues shown as sticks have been colored blue regardless if they are in domain I or III. Image created in PyMOL[18]

The bonds theorized between Her2 (3WLW) and EGF (1EGF)

Asn1, mainchain NH group – Gly324, mainchain CO group (Hydrogen bond)
Asn1, sidechain CO group – Gly324, mainchain NH group (Hydrogen bond)
Ser2, mainchain CO group – Arg12, 2nd/3rd sidechain amino group (Hydrogen bond)
Cys6, mainchain CO group - Arg332, 2nd/3rd sidechain amino group (Hydrogen bond)
Pro7, mainchain NH group – Glu357, sidechain COOH group (Hydrogen bond)
Ser8, sidechain hydroxyl group – Arg332, 2nd/3rd sidechain amino group (Hydrogen bond)
Glu24, sidechain COOH group – Gln424, sidechain NH2 group (Hydrogen bond)
Asp40, sidechain COOH group – Leu101, mainchain NH group (Hydrogen bond)
Arg41, 2nd and 3rd sidechain amino group – His66, sidechain amino group (Hydrophilic interaction)
Leu47, mainchain CO group – Thr45, mainchain NH group (Hydrogen bond)
Arg53, mainchain NH group – Glu18, sidechain COOH group (Hydrogen bond)
Arg53, mainchain COOH group* – His448, sidechain amino group (Hydrogen bond)
Arg53, sidechain amino groups – His448, sidechain amino group (Hydrophilic interaction)

*Arg53 is the C-terminal residue and therefore has a COOH group in its mainchain as it is not involved in a peptide bond.

In figure 19, EGF is shown docked in the pocket between domains I and III of Her2, with the theorized bonds included to show the strength of the binding.

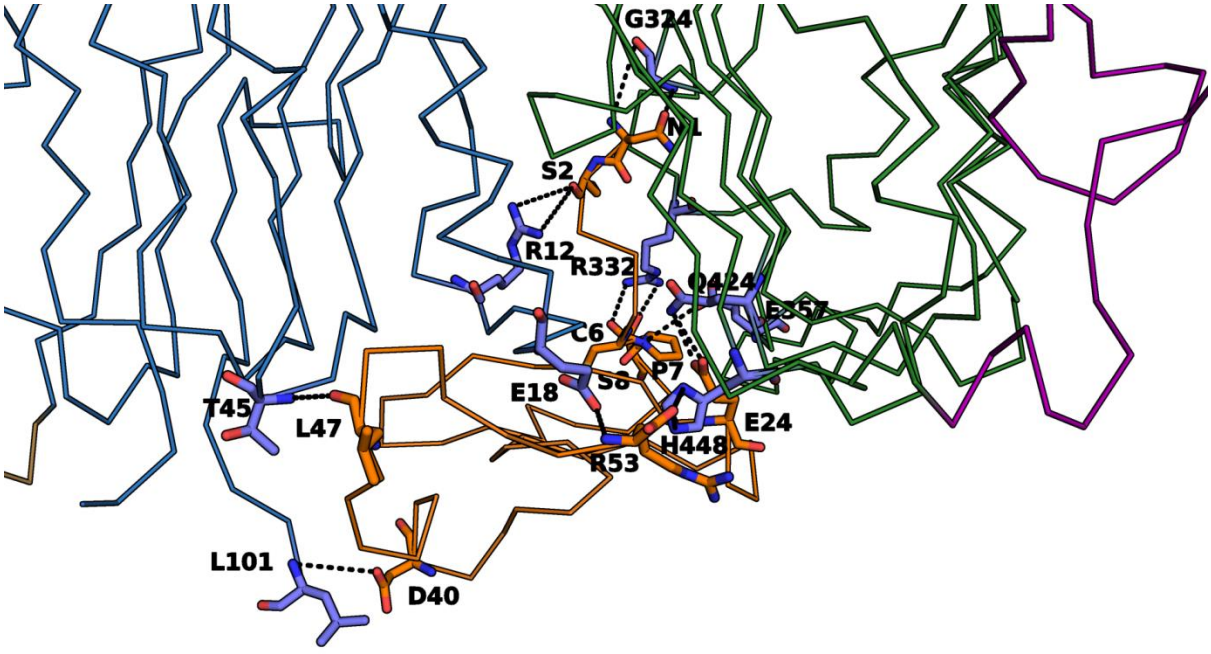

Figure 19: EGF (orange) docked between domains I (blue) and III (green) of Her2. Theorized bindings are shown with dashes between the residues, which are shown in sticks, note that all Her2 residues shown as sticks have been colored blue regardless if they are in domain I or III. Image created in PyMOL[18]

Docking result from gridbox 9

AutoDock Vina found 9 docked structures for both EGF and TGF-α, all with negative binding affinities. The values of the binding affinities can be viewed in table 9 and 10 in appendix 3. The best resulting structures of both ligands were inspected and both found to be docked in such a manner that they enveloped the dimerization arm (residues 245-266).

For TGF-α the residues that is close to the dimerization arm is 5-9, 22-27, 43-50. With the first two groups being on the side of domain IV and the residues 43-50 wrapping around the dimerization arm to the other side.

The residues of docked EGF that are close to the dimerization arm is residues 1-5,8,11,12, 18-21 on the side of domain IV, with 47-53 wrapping around the backside of the dimerization arm.

Based on the observations above and a closer inspection the following bindings were suggested:

The bonds theorized between Her2 (3WLW) and TGF-α (2TGF)

Asp7, mainchain NH group – Thr251, sidechain hydroxyl group (Hydrogen bond)
Arg22, sidechain 1st amino group – Thr254, sidechain hydroxyl group (Hydrogen bond)
Phe23, mainchain CO group – Asn253, sidechain NH2 group (Hydrogen bond)
Gln26, sidechain NH2 group – Thr256, sidechain hydroxyl group (Hydrogen bond)
Gln26, sidechain NH2 group – Asn253, sidechain CO group (Hydrogen bond)
Glu44, sidechain COOH group – Tyr252, mainchain NH group (Hydrogen bond)
Ala50, mainchain NH group – Tyr252, sidechain hydroxyl group (Hydrogen bond)
Ala50, mainchain NH group – Val250, mainchain CO group (Hydrogen bond)

In figure 20, TGF-α is shown docked to the dimerization arm that is found in domain II of Her2. The theorized bindings between the dimerization arm and TGF-α is shown in the figure. Note that the dimerization arm is centered in the figure, and the rest of domain II is below the dimerization arm, outside the view of the figure.

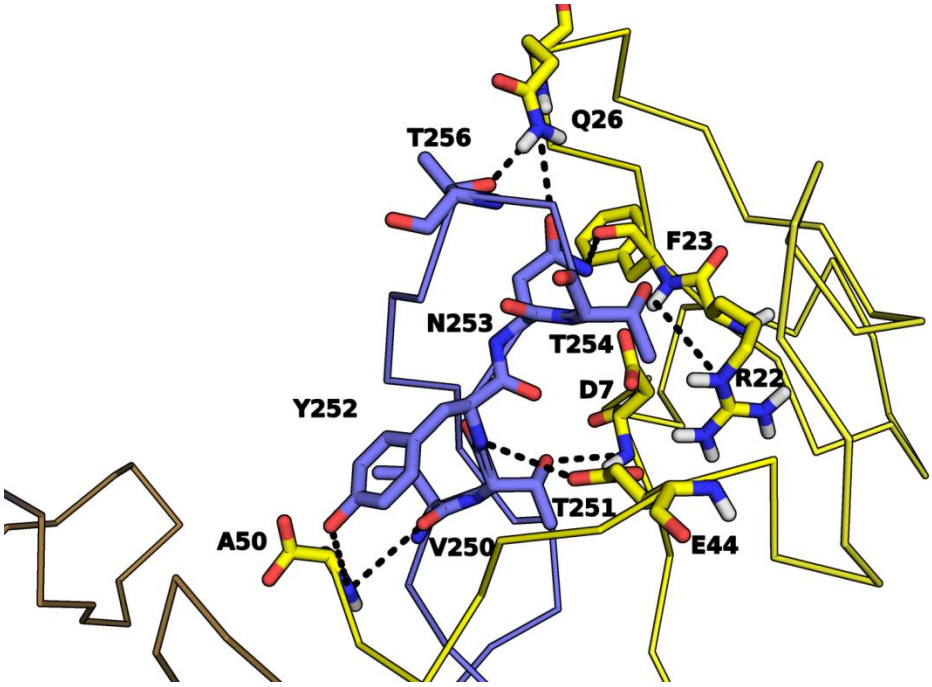

Figure 20: TGF-α (yellow) docked to the dimerization arm (blue) in domain II (brown) of Her2. The residues involved in bindings are shown in "sticks", with dashes between them representing different type of interactions and bonds. Residues are labeled using the one letter name for amino acids. Image created in PyMOL[18]

The bonds theorized between Her2 (3WLW) and EGF (1EGF)

Asn1, sidechain NH2 group – Met260, mainchain CO group (Hydrogen bond)
Asn1, mainchain NH group – Ser259, mainchain CO group (Hydrogen bond)
Gly12, mainchain NH group – Glu264, sidechain COOH group (Hydrogen bond)
Trp49, sidechain amino group – Thr254, mainchain CO group (Hydrogen bond)
Trp50, sidechain - Phe257, sidechain (Hydrophobic interaction)
Arg53, $2^{nd}/3^{rd}$ sidechain amino group – Ser259, sidechain hydroxyl group (Hydrogen bond)
Arg53, $2^{nd}/3^{rd}$ sidechain amino group – Phe257, mainchain CO group (Hydrogen bond)
Arg53, mainchain COOH group* – Glu258, mainchain NH group (Hydrogen bond)
*Arg53 is the C-terminal residue and therefore has a COOH group in its mainchain as it is not involved in a peptide bond.

In figure 21, EGF is shown docked to the dimerization arm that is found in domain II of Her2. The theorized bindings between the dimerization arm and TGF-α is shown in the figure. The dimerization arm is centered in the figure, and the rest of domain II is below the dimerization arm, outside the view of the figure.

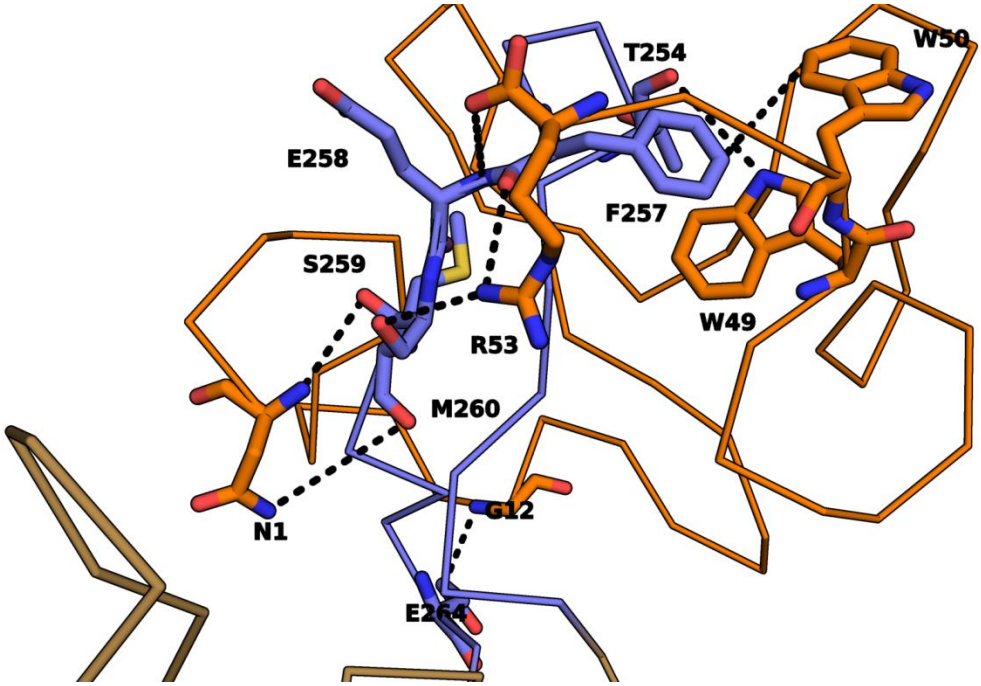

Figure 21: EGF (orange) docked to the dimerization arm (blue) in domain II (brown) of Her2. The residues involved in bindings are shown in "sticks", with dashes between them representing different type of interactions and bonds. Residues are labeled using the one letter name for amino acids. Image created in PyMOL

Summation of AutoDock Vina results

Table 1: Summary of all the Vina docking experiments. Showing the binding affinity of the best result from each docking experiment. For gridboxes 1, 2,3,4,5 and 6 v1 the docking experiments was done with only TGF-α as ligand, and the EGF cells are therefore blacked out to represent that no experiment was done. The dash represents that there was done an experiment but no results were found.

	TGF-α		**EGF**	
Gridbox	Number of docked structures	Binding affinity of result #1 (Kcal/mol)	Number of docked structures	Binding affinity of result #1 (Kcal/mol)
1	0	-		
2	2	435,4		
3	0	-		
4	0	-		
5	0	-		
6 v1	1	277,8		
6 v2	9	-3,0	9	-2,2
7	9	-3,5	9	-2,1
8	9	-4,9	9	-4,6
9	9	-3,7	9	-3,2

In summation there was found four different dockings for both TGF-α and EGF. All of these four bindings had negative binding affinites, and therefore many possibilities for bindings and interactions. The results from all of the docking experiments, both the ones with good results and bad results have been summed up in table 1. It serves as an overview over which gridbox had the best binding affinity as well as the number of structures that was found by AutoDock Vina.

Ramachandran plot of the docked ligand structures

A Ramachandran plot was made for each of the docked structure (the first conformation only), except the two docked structures that yielded positive binding affinities. From each of the Ramachandran plots, the percentages of residues in favored and allowed regions was determined and are found in table 2. A Ramachandran plot was also done of the EGF and TGF-α structure from the pdb database which was the structures used in the AutoDock Vina experiments. These can be used for comparing with the structures that has been docked to Her2.

Table 2: Table showing percentage of residues in favored regions and allowed regions of the different structures from AutoDock Vina experiments. Vina run 6 v2 was the docking experiment done with gridbox 6 v2 etc. The Ramachandran plots were made using the MolProbity tool [39].

Source	**TGF-α**		**EGF**	
	% residues in favored regions	% residues in allowed regions	% residues in favored regions	% residues in allowed regions
Protein database	64,6	89,6	62,7	88,2
Vina run 6 v2	58,3	91,7	63	96,3
Vina run 7	64,6	95,8	63	92,6
Vina run 8	58,3	89,6	63	92,6
Vina run 9	66,7	89,6	66,7	92,6

Docking TGF-α to other receptors/proteins

Results were found for both search areas made for the test protein 1WP6, 9 docked structures were found for both search areas. The complete list of binding affinities of TGF-α to 1WPL6 can be found in tables 11 and 12 in appendix 3. The negative values of the top results from both search areas, means that AutoDock Vina found a binding possibility between TGF-α and 1WP6. These can be compared to the values of the dockings of TGF-α to Her2 which are summarized in table 1 on page 45.

The specifics about which residues are involved in the binding in these two spots are of no interest as the only data wanted from this “control” experiment is the binding affinity and whether the binding is sterically possible. Nor is a figure of the structure, but it can be noted that the docked structure was a good docking and contained no steric problems.

Discussion

Comparing structures of Her2

There were three structures of Her2 compared to each other, 3WLW, 1N8Y and 2A91 (see page 16). Based on the structural comparisons it can be seen that all three are very similar structurally. 3WLW and 2A91 were near identical in the first three domains, but 2A91 had its domain IV much shorter. 3WLW, while still very similar, was less identical to 1N8Y, but 1N8Y had a larger domain IV. Due to the shortened domain IV, 2A91 can not be used in some of the docking experiments as domain IV is of much interest. 1N8Y is a structure of rat Her2 (Neu) and therefore has some differences, albeit small, from human Her2. Despite having a longer domain IV, 1N8Y will not be used as the main Her2 structure due to being a structure of Neu rather than Her2. 3WLW will therefore be used as the main structure for molecular docking in this thesis. If some of 3WLWs attributes makes it not suitable for a specific experiment, the most suitable structure will be chosen.

Comparing the binding possibilities of Her1 and Her2 based on structural and sequence alignments

Is the DALI structural alignment of good quality?

From the DALI alignment there were several output structures made by the DALI server, the first output was chosen for further use. Assessing the DALI alignments can be done by looking at calculated values that represents the quality such as the Z-score or the RMSD but can also be done by looking at the structure and manually assessing it. For choosing which of the alignments was to be used, the Z-score was the most important and therefore the result with highest Z-score was chosen, even though it had a higher RMSD than the results immediately ranked lower on the Z-score. Despite the alignment with highest Z-score was favored to be the one to use for further analysis, a manual inspection of the top ranking results (ranked by Z-score), was also performed.

In the analysis the result with the highest Z-score was found to have a nice global alignment, where the whole structure of both proteins had been super positioned over each other. This alignment is depicted in figure 11. The alignment was also done in such a way that the corresponding domains of both Her1 and Her2 were on top of each other so the alignment was as close as could be (This is depicted in figure 11 and figure 22 A) . This compared to other structures such as the one with second highest Z-score (shown in figure 22 B, and its RMSD and Z-score is listed in Appendix 1) which had only domain III of Her1 on top of

domain I of Her2, and the rest of the proteins where not aligned to each other. While this alignment had a lower RMSD, the result is not a good alignment since it only achieves a local alignment of one part of each of the protein, instead of a global alignment. The lower RMSD therefore is only a result of placing two similar (even though it was two different domains from Her1 and Her2, remember that both domain I and III are L domains and are very similar) structures on top of each other while ignoring the rest of the proteins. As only the distance of the aligned atoms are included in the RMSD calculations the RMSD becomes low. Overall the alignment from DALI that has been chosen to be used for further analysis is a good alignment between two proteins that shows their similarity by placing the corresponding domains from each protein on top of each other. Furthermore the Z-score has a very high value, and the RMSD is not high enough to be of large concern. Thus the resulting alignment from the DALI server is a good representation of the similarities in structure between the Her1 structure 1IVOa and the Her2 structure 3WLWa.

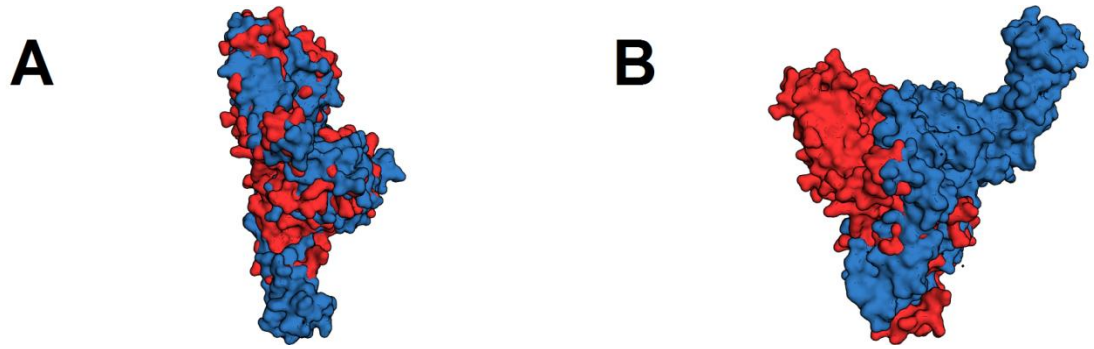

Figure 22: A) Showing the surface of 1IVO (red) and 3WLW (blue) aligned, based on the top resulting structure from the DALI alignment**. B)** Showing the surface of 1IVO (red) and 3WLW (blue) aligned, from the second best structure from the DALI alignment. In both A) and B), domain I of 1IVO is on the top part of the picture, while domain IV is towards the bottom, Her2 is viewed from above compared to the other depictions of the Her2 ectodomain in this thesis. Image created in PyMOL [18]

From the results from the DALI alignment between 1IVO and 3WLW (i.e. an alignment between Her1 and Her2), there was also given a sequence alignment that accounts for the structure similarity in addition to sequence similarity. This sequence alignment will be used in conjunction with the information about interactions and bonds given by Ogiso et al. (2002) and Garret et al. (2002) to assess if there's a possibility of a binding between Her2 and TGF-α/EGF similar to the one found between Her1 and the two ligands.

Her1 EGF binding

Between the binding sites of Her1 and Her2 there is a high similarity in structure, but there is not the same level of similarity in the amino acids composition between the binding sites of Her1 and the corresponding sites in Her2. While site 1 is relatively well conserved, site 2 and 3 is less so. This can be seen in the first of the sequence alignments in Appendix 1, where site 1 have a relatively good match, while site 2 and 3 have been aligned badly. Site 2 and 3 does not have a good sequence alignment in any of the other sequence alignments from the DALI run.

Site 1, which is found in domain I, is seemingly well preserved. When comparing the structure and the amino acids it can be seen that there is a great similarity between this area of Her1 and Her2. When looking at the amino acids that are involved in interactions and bindings with EGF there are also similarities. Of the 4 residues involved in hydrophobic interactions with EGF only one is different between Her1 and Her2, the Leu69 of Her1 corresponds to His66 in Her2. The bindings between Her1 occurs in a hydrogen bond for Glu90 , which is reserved in Her2, and a beta-sheet like binding involving three residues in both Her1 (16-18) and EGF(31-33) [16]. The corresponding residues in Her2 has been mutated, Gln16 →Leu13, Leu17→Pro14 and Gly18→Ala15. These mutations should not matter greatly as the beta-sheet like bindings does not involve sidechains. Overall the difference in amino acids involved in the site should not be too big of a hinder for binding in her2.

In site 2 (Domain III) two of three residues involved in interactions with EGF have mutated, but the structure is well conserved. The two residues involved in hydrophobic interactions, Val350 and Phe357, have mutated to Glu357 and Ala364 respectively. The ability of making hydrophobic interactions is affected by the Val→Glu mutation. The salt bridge that Asp355 is involved with, which is the strongest bond between Her1 and EGF in site 2, could happen in Her2 as well because an aspartate residue (Asp362) is found in a similar position. Overall in spite of the mutations, there could be a binding in this site of Her2 due to the well conserved structure and the possibility of 1 hydrophobic interaction and a salt bridge.

Site 3 (Domain III) is the least conserved site in concern to mutations of residues involved in bindings and interactions. Leu382, Phe412 and Ile 438 are all involved in hydrophobic interactions between Her1 and EGF. In Her2 they have been mutated to Tyr389, Tyr419 and Leu445 respectively. While the tyrosine does have a positive charge, depending on how the structure is, the hydrophobic parts of the sidechain of tyrosine could allow for hydrophobic interactions and the Leucine, which is hydrophobic, would of course be able to participate in hydrophobic interactions. The side chain of Gln384 in Her1 has a hydrogen bonding to EGF, in Her2 the corresponding residue is Ser391, whose OH group may be involved in a similar binding.

Her1 TGF-α binding

The pdb structure 1MOX is a structure of a TGF-α:Her1 complex dimer[17]. The structure shows the binding of TGF-α to EGFR/Her1 as well as the dimer between two TGF-α:EGFR complexes. The bindings between TGF-α and EGFR is described in the article that goes with the structure by Garret et al.[17]. The article describes specific interactions between different residues in both TGF-α and EGFR. The comparison of these bindings with potential bindings with Her2 will be based on a structural alignment done with a different structure of Her1 (1IVO) than the structure that the information about the TGF-α:Her1 complex is based on. This does not matter as the two structures are of the same protein and are identical in sequence (see figure 23) and should be almost identical in structure.

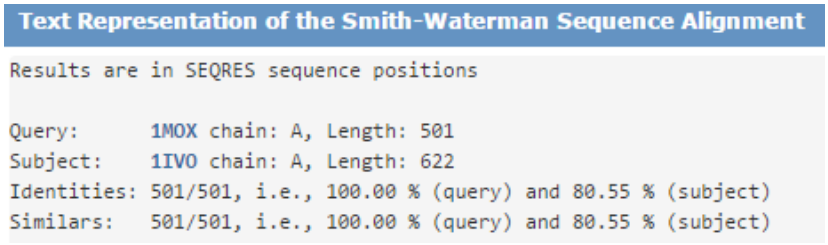

Figure 23: Screenshot of a sequence alignment of the sequences of protein structures 1MOXa and 1IVOa, two different structures of Her1.
Sequence alignment done by http://www.rcsb.org/pdb/workbench/workbench.do tool.

In domain I there are 11 residues involved in some form with the binding of TGF-α, which corresponds to "site 1" used when explaining the binding between Her1 and EGF. The residues 15-17 of Her1 binds to TGF-α, with the residues forming a β-sheet like interaction with residues 29-35 of TGF-α. In Her2 none of these residues are conserved but it is not of importance as the β-sheet bindings is between the main chain atoms and not sidechain, and should therefore be able to have a similar binding. Asn12, which has its side chain involved in a binding/contact with the main chain N of Gly40 in TGF-α is replaced with Met9 in Her2 and is therefore unable to make such a contact. Thr15 in Her1 forms a hydrogen bond, in Her2 the Threonine residue has been exchanged with Arg12, and it therefore has no oxygen in the side chain to make a similar hydrogen bond. The two leucines, Leu14 and Leu17 are both involved in hydrophobic interactions Leu 14 is conserved in Her2 (Leu11) while Leu17 corresponds to Pro14, which means both residues have the ability to have hydrophobic interactions with TGF-α. Regardless of the very important changes that could prevent binding (Asn12→Met9 and Thr15→Arg12 are especially crucial for the ability to bind TGF-α [17]), a docking experiment should be done as it could find potential for binding.

The residues involved in binding are less conserved in the domain III than the domain I, but their nature and ability to make similar interactions are higher than in domain I. Asp355 which forms a salt bridge with TGF-α in Her1 is conserved in Her2(Asp362); this means that the only bond between the L2 domain and TGF-α is present in Her2. There is also a reserved residue that is involved in hydrophobic interactions; Leu348 (corresponds to Leu355 in Her2). Most of the other interactions are done by residues involved in hydrophobic and hydrophilic interactions. They have been replaced with residues that in most cases have different charge than their counterpart in Her1 and therefore have lost the ability to have an interaction of the same kind. These changes are Leu325→Arg332, Phe357→Ala364, His346→Ala353 and Val350→Glu357. While others have also been replaced but kept their polarity and can have hydrophilic interactions, these changes being, Gln384→Ser391. His409→Asn416 and Gln408→His415. There is also a hydrophobic pocket in the L2 domain created by residues that holds Leu49 of TGF-α buried in it and it is the residue with highest surface area buried of TGF-α. When comparing these areas between a Her1 and Her2 structure it can be seen that a similar hydrophobic pocket exists in the same area. There are many similarities between the binding sites for TGF-α in Her1 and the tentative binding site in Her2, but there are unfortunately more disparities than similarities.

Closing remarks about the comparison of Her1s binding and Her2s tenative binding

If the assumptions and pairwise sequence alignment is wrong, and wrong Her2 residues are said to be the equivalent of the Her1 residues involved in binding of Her1, it should not matter greatly in regards to results from Vina. As the gridboxes covers a larger area that is structural equivalent of her1s binding residues positions. If the chosen residues are wrong, and there are other residues that are the "real" equivalents they will be inside the search area (as the search area still covers all the structure equivalent to the structure involved in binding between Her1 and EGF/TGF-α), and a successful result will still be possible.

In spite of the differences between Her1 and Her2, comparing them and finding similarities is a good method for finding a potential binding site. Throughout evolution of proteins,the structure is always more conserved than the sequence. Her1 and Her2 are closely related and has the same origin, their structure is therefore very similar, and the major difference is in the composition of amino acids and not in the structure itself. Because of this structural similarity and amino acid composition disparity, a comparison of the two structures have given great insight into Her2s possibility of binding EGF/TGF-α. Despite the similarities not being overwhelmingly high, they are still at a level where Her2 can be considered to be similar enough to Her1 to have similar bindings. The structural alignment, and comparison of Her1 and Her2, did not give enough information to justify concluding if a Her2 can bind either of the ligands similar to Her1.

In order to further check for bindings, docking experiments will have to be done and the results analyzed.

Evaluation of the docked structures

A general discussion of methods and results

A study indicates that there is an optimal size for search areas, and proposes a formula for calculating the optimal size. If this is accurate, the search areas that were used to get the results in this thesis were not of optimal size as they were not calculated by the aforementioned formula. According to Feinstein & Brylinski [50] using a too small gridbox would not yield enough good results while using one that is too large would give many irrelevant bindings that are not necessarily the ones you search for. They further suggest a method of finding the optimal size of a search area using a formula they made based on a large study.

While the size of the search area certainly could be a problem, the issue of the search area is not comparable between the two cases. They do their research by using the search areas in virtual screening (docking a large number of different ligands to a receptor of great interest) where they are docking many different ligands to a certain area. In the case of virtual

screening you have a pre-conceived area where you want your ligands to be bound, e.g. looking for a ligand that binds to a binding site. In the case of the search areas for Her2, the search areas are heavily based on other bindings and certain residues are of great interest, but there is no special area where the binding has to be found for the experiment to be successful. Therefore these residues have been included in the search area so that the ligands have a possibility of binding these ligands. The optimization of the gridboxes by Feinstein & Brylinski is a procedure that makes the search area large enough to get all possibilities of binding in the site of interest but ignores all dockings that would happen outside of the area of interest. The search areas used in this thesis is meant to give the docking experiments the possibility of searching for a binding that mimics that of known Her2 ligands, but still be open for possibilities of a strong binding that is not completely in the binding site. Therefore such optimization of the search area is not relevant or useful for the work in this thesis.

Furthermore there could be an issue of the search area not being large enough to allow for good rotations of the ligand (as EGF and TGF-α are large), such as the search area calculated by Feinstein & Brylinskis method would allow. In order to make sure that the ligand would have space for itself and to have the possibility of rotation the ligand was manually docked to the receptor, by moving the coordinates of the ligand using AutoDock Tools. The gridbox was then made around the docked ligand (including the wanted residues as explained above and earlier) in such a manner that the ligand was allowed full rotation inside the search area. The gridboxes used in this thesis (the ones used to get the results, not the first iteration with poor results), had the size that was needed for them to explore all options and get optimal dockings within the areas of interest.

Regarding the binding affinities that are given alongside the docked structures they are uniform for the successful dockings. All of the results from the successful dockings of gridboxes 6 (v2), 7, 8 and 9 have high negative binding affinity values; this indicates that the bindings are strong. This opposed to the binding affinities of the wrongful or false positive dockings (gridbox 2 and gridbox 6 v1) where the values for the binding affinities were high positive values. There was a great disparity between the values of the successful dockings and the false positive dockings E.g. the binding affinity of the best result from gridbox 2 was 435,4 kcal/mol compared to the binding energy for gridbox 6 v2's best result that had a binding affinity of -3,0 kcal/mol (for docking of TGF-α). The extremely high values of the result of gridbox 2 shows that a binding between the ligand and receptor would not be thermodynamically favorable, and would therefore not happen. For the other successful bindings of both EGF and TGF-α the negative binding affinities shows that the bindings are thermodynamically favorable and are bindings that could take place in a living cell.

Is the binding of EGF/TGF-α to Her2 in Vina run 6 v2 a good binding?

The data from the docking experiments with gridbox 6 v2 as the search area indicates a possibility of binding of EGF and TGF-α towards the end of Her2s domain IV.

If this binding is indeed correct it would be very interesting that TGF-α and EGF binds to the same area and residues that Herceptin binds. What this means can only be theorized without doing further *in vivo* or *in vitro* research on Her2 systems.

It could be that Herceptin's mechanism of action involves blocking EGF/TGF-α from binding to Her2 by being a competitive inhibitor, which would explain Herceptin's mode of action but not TGF-αs or EGFs mode of actions. The binding of TGF-α or EGF could at some point in time, before Her2 evolved to not having an inactive form (where the dimerization arm is bound to domain IV), could act as a competitive inhibitor of the formation of the inactive form. Summing up, one theory for explaining this binding is that EGF/TGF-α binds to domain IV and the binding has an unknown function that Herceptin hinders when it binds as a competitive inhibitor.

Another impact of the binding of TGF-α and EGF to the same space as Herceptin is that they could have similar effects to Herceptin when it comes to blocking metalloproteases from cleaving off Her2s ectodomain [46]. As Her2 is always in its active configuration, domain IV remains open for ligands to bind to it and enzymes to act on it. Therefore there is a theoretical possibility of the metalloproteases cleaving at the unprotected domain IV, removing the ectodomain from Her2 and making the kinase domains active. This could have disastrous implications on the signaling systems of Her2. This may be hindered by EGF or TGF-α binding to domain IV and acting as a steric blocker in a similar mode as Herceptin. The binding of EGF/TGF-α could be to protect, against proteases cutting the ectodomain of the rest of the Her2 protein.

Herceptin has more modes of action, one of which causes Her2s rate of endocytosis and the subsequent degradation of Her2, to be increased. The exact mechanics of this process is not known, but it is known that the binding of Herceptin promotes the rate of the process. It is possible that the binding of EGF/TGF-α in the same area as Herceptin, could have similar effects on the endocytosis rate of Her2 as Herceptin has. This could potentially be some type of regulating of the signaling system of the Her family. EGF/TGF-α binds first to Her1 and induces dimerization with other Her1 proteins but also Her2 proteins. It could be that, after a dimerization has happened between Her1 and Her2, EGF/TGF-α will bind to domain IV of Her2 and induce endocytosis in order to stop the signaling of the dimer.

The number of theorized bonds indicates that the binding is strong. There is a good amount of theorized bonds to Her2 for both TGF-α and EGF. In addition to the bindings showing the strength of the binding some of them show the similarity between Her2s binding of Herceptin

and the docked structure of the ligands. There are several residues of Her2 that participates in the binding of Herceptin that also participates in the binding of the ligands. Such as residues 572 and 573 being part of the second binding site of Herceptin[28] as well as them seemingly being involved in hydrogen bonds in the docked structure of Her2s domain IV and TGF-α. The docking between TGF-α/EGF and Her2 does not only take place in the same area as the binding of Herceptin to Her2 but does also include bindings of some of the same residues.

The docked TGF-α and EGF is not involved in any form with the residues that corresponds to the residues involved in the tether between domain II and IV in Her1. Nevertheless the docked ligands are close enough to the tether site that there could be some steric hindrance from the docked ligands that influences interactions that would happen with the residues involved in the tether.

Taking everything into consideration, the binding of both EGF and TGF-α to domain IV of Her2 is a strong binding. It has a large negative binding affinity as well as many possibilities of bonds between the ligands and the receptor. In addition the binding of either ligand may have impacts on Her2 similar to the effects of the binding of Herceptin, or alternatively have another important impact on Her2. The binding of the two ligands, EGF and TGF-α to domain IV of Her2 are both strong bindings with the possibility of the binding having an effect on Her2.

Is gridbox 7 a reasonable binding, would the binding have any impact on the role of Her2?

There are no direct implications that there would be an effect of the binding of EGF or TGF-α to domain II of Her2. The binding of IgG fab did not have any effect of the ability to homo dimerize according to results and structure from Hu et al.[12] (Though it did have a head to tail binding between the two Her2s opposed to the more accepted tail to tail binding). Therefore there are no reasons to believe the binding of the ligands in the same position would affect dimerization either. Thus the binding does seem a little suspicious as there is no theoretical effect on Her2 caused by the binding. Of course not having a theory on what effect the binding would have on Her2 does not directly indicate that the binding is false. Nevertheless it does give a reason to be more cautious of the results for these two bindings and more careful with their interpretation.

Regardless of the suspicions described above, there are many theorized bonds for the two dockings. That fact combined with a negative binding affinity indicates that a binding between the ligands and receptor in this area is possible.

Is the binding between domain I and III (gridbox 8) comparable to its Her1 counterpart and would the binding of ligand have similar impact?

From the docked TGF-α structure with largest affinity, there are many similarities between the docked structure and the TGF-α observed in the TGF-α-Her1 complex. Most of these similarities can be seen in how the ligand lays in the gap between domain I and III with the N terminal and C terminal in same ends as in the Her1 structure. In addition the rest of the protein is orientated in a similar fashion as the TGF-α which is bound to Her1.

Regardless of the docked structures lack of similarity to the structure of Her1, which had been theorized by comparing residues and therefore anticipated, there is still great potential of binding.

Despite the many possible hydrogen bonds and hydrophobic interactions between both ligands and Her2, they are vastly different from the bonds that was suggested based on the structural and sequence alignment between Her1 and Her2. In the case of EGF only two residues of the ones suggested to have a binding, have been theorized to have a binding in the docked residues. These two are His66 and Glu357, but both of these residues have different type of interactions than their Her1 counterparts. As His66 is involved in a hydrophilic interaction and not a hydrophobic interaction and Glu357 have been theorized to be involved in a hydrogen bond while its Her1 counterpart was involved in a hydrophobic interaction. For TGF-α the disparity between the theorized bonds in the docked structure and the bonds that was suggested bonds based on the alignments is also large. Only one residue that was suggested from the alignments have taken place in the docked structure of TGF-α and Her2, that residue being His12 being involved in an hydrogen bond, just as its counterpart in Her1 also participated in. In neither of the structures was the hydrophobic pocket involved in the binding, even though it was well conserved in Her2. There was residues buried in a pocket, but that was the N-terminal of EGF buried in another pocket that was of a more hydrophilic nature. Overall there is very little similarity between the binding of EGF and TGF-α to Her2 that has been observed in the docked structure, and the theorized binding based on the Her1 binding of the ligands.

While there are many theorized bonds and thus a strong binding, the difference from the suggested bonds is a problem. It does not appear that Her2 is able to bind EGF and TGF-α in a similar manner to Her1. This definitely raises questions to the validity of the docking in the cleft between domain I and III. As it's not similar to the one in Her1, the binding can't be a binding that has been kept through Her2s evolution from Her1. There does not seem to be any direct implication of the binding either, as the binding cannot release the dimerization arm from domain IV as no such binding exists in Her2 and it cannot change the structure of domain II either as domain II of Her2 does not change structure. The binding is not a "relic" from times past, as it's different to the ligands' binding to Her1 and if the binding takes place,

it is so different that it must have been developed independently. The binding does not have any apparent function; this weakens the theory of this binding taking place.

Nevertheless there is good data supporting the binding for both the ligands with a large negative binding affinity and a good number of bonds and interactions being possible, with the chance of more not being noticed. Therefore the docking results cannot be put aside and the binding of the two ligands EGF and TGF-α to the cleft between domain I and III in Her2, has to be viewed as a possible binding of EGF/ TGF-α to Her2.

Gridbox 9 , does EGF or TGF-α bind to the dimerization arm in a similar manner to Perjeta?

Both EGF and TGF-α was found to dock to Her2 by AutoDock Vina. The bindings of the two different ligands have some similarities but also some differences. Neither of the bindings are completely identical to the one of Perjeta, as where Perjeta binds completely on one side of the dimerization arm, EGF and TGF-α's binding is more enveloped around it. At the same time most of the ligands are on the opposite side of the dimerization arm than Perjeta. Even though the TGF-α/EGF binding does not cover the same areas the Perjeta binding it still envelops and covers the dimerization arm in such a manner that it sterically blocks it from interactions with other molecules and therefore could stop dimerization and therefore have a similar effect to Perjeta.

If the two ligands binds to the dimerization arm this would stop the potential of Her2 dimerizing, as the dimerization arm can not interact with the dimerization arm of the dimerization partner. This seems to be a very counterintuitive ability of Her2. Why should the ligands that initiates the dimerization process in Her1, stop the dimerization process in Her2, which is a potential dimerization partner for Her1. Of course the data can not be trivialized just due to this seemingly double agent role of the ligands. There was found many theorized bindings, and the binding affinity was negative, which indicates a binding. The binding is definitely possible but the most obvious impact of the bindings goes against what one would expect the role of the ligands to be, and makes the binding seems less likely. Nevertheless the data from AutoDock Vina does indicate that there is a definitive binding potential for both EGF and TGF-α to the loops in the beginning of domain II.

Analyzing the impact of AutoDock Vina's "flex" on the EGF and TGF-α structures

By doing a Ramachandran plot of the output from the Docking experiments it is possible to check if the structures have been changed by the flexibility feature of AutoDock Vina, and check if the changes were unfavorable. In table 2 on page 46, the percentages of residues in favorable and allowed regions are showed. From the table it can be clearly seen that all of the modified structures from the output ligands from Gridboxes 6, 7, 8 and 9 have better or similar percentage of residues in allowed regions. This indicates that the sum of changes in the structures is positive, as the percentage of residues in allowed regions does not go down. Thus the structure outputs from Vina are structures that can exist without any steric or stability problems caused by the torsion angles.

Some of the docking result structures have a higher percentage of residues in favored areas than the pdb structures for EGF and TGF. This does not necessarily mean that the changed structures are more stable or better. Regardless of that, the higher percentage of residues in favorable areas does bolster the claim that the structures are good structures that could be close to conformations of the ligands yet to be observed.

Reflections on the results from the "blind test"

The results of the blind test of docking TGF-α to the amylase (pdb: 1WP6), an α/β protein with 485 residues, were both positive. There were found at least 9 structures for both gridboxes used, where there were at least 5 structures for both search areas with negative binding affinity. For one of them they were overwhelmingly so with the binding affinity indicating a stronger binding than for some of the found bindings of TGF-α to Her2.

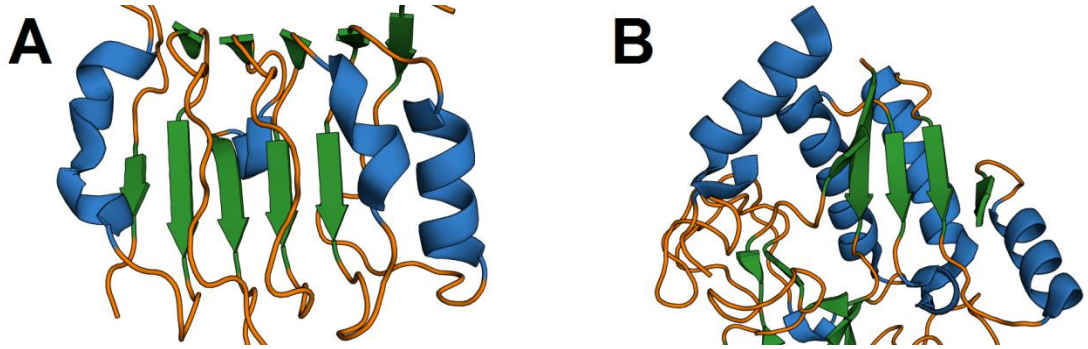

Figure 24: A) domain I of Her2 shown in cartoon. **B)** A cross section of the structure 1WP6 (an amylase structure).The structures are colored by secondary structures,α-helix blue, β-sheets green and loops orange. Image created in PyMOL[18]

Coincidentally the architecture of the secondary structure of the test receptor (1WP6) resembles the architecture of L domains of the Her family. The amylase 1WP6 is a protein that consists of a domain/architecture/topology that is commonly known as a TIM barrel or β-barrel. A β-barrel consists of 8 centered β-strands that form a ring or a barrel surrounded by

different secondary structures that are there to support the β-sheets that form the barrel[31]. The L domain of the Her family have 4 β-strands central in the domain with different secondary structure around it, which is shown in figure 24. These 4 β-strands are very similar to a half β-barrel, in addition more β-sheets are found outside of the barrel structure that is also very similar to the β-sheets from the L domain of the Her family. This similarity is shown in figure 24, where domain 1 of Her2 is on the left, and the amylase's β-barrel has been cut in half and is shown on the right. In the architecture of a TIM barrel each beta strand that makes up the barrel is preceded or/and followed by an α-helix, this is also the case for some but not all of the β-sheets in the L domains. While the similarity is not high enough to call the two different architectures/topologies identical, they are very similar. This could mean that some of the reason for the surprising bindings of TGF-α to the amylase could be explained by the similarity between the domain found in Her2 and the amylase.

The binding can be an indicator of a problem with the docking experiments. TGF-α (and most likely EGF as well) seems to have a tendency to bind to a protein when docked with AutoDock Vina. Where the problem comes from is hard to figure out but it could be TGF-α being a protein that has many surface residues that can easily participate in hydrophobic and hydrophilic interactions. The problem could also stem from AutoDock Vina's formula being biased in such a way that when using proteins of a size such as TGF-α as ligands, they will always easily be docked to a receptor. Alternatively the problems could be related to AutoDock Vina but rather than being the fault of the formula it could be that with search areas this large it will easily find a docking. There is no way to assess which of the suggested problems is the real fault, nor is there a way to circumvent them short of changing docking program in hope of AutoDock Vina being the problem. Changing the ligand would of course be of no use and change of the search areas by increasing or decreasing their size is not an option, with reasons for this stated earlier in the discussion.

How much this impacts the results from the other docking experiments is hard to tell. Even though the new results may lead to thoughts of the results achieved from docking TGF-α and EGF to Her2 being false positives, they may not necessarily be the case. While these dockings do show that the docking software will dock TGF-α to proteins that it should not bind to, this does not disprove all of the results from the docking experiments. First it has to be considered that there has been observed other bindings to the areas where TGF-α and EGF have been docked onto Her2 have been observed, so there are a definitive possibilities of binding in those areas. Furthermore bindings such as the one in the pocket between domain I and III of Her2 have similarities to the binding of EGF and TGF-α to Her1 and therefore have credibility from this. In addition the possibilities of interactions have been discussed, and thereby giving and explanation of why the bindings are possible. This gives the docking results (between Her2 and TGF-α/EGF) credibility.

Even though one of the two test dockings had good binding affinities, it does not invalidate the experiments. The similarities between Her2 and the amylase together with the strong theoretical reasons for why the bindings are possible, gives the docking result credibility and indicates authenticity.

Comparison of the different docked structures

If there exists a binding between either of the ligands and Her2, there is most likely only one binding and not several, despite there being found several bindings from the docking experiments. Therefore the bindings will be compared in order to try and determine which one is the most likely binding.

Even though the binding affinity is highest for both TGF-αs and EGFs docking to Her2 in gridbox 8, does this fairly reflect the binding? The binding affinity given by AutoDock Vina describes the total energy of the bonds and interactions that takes place between the docked ligand and receptor. While one would assume that the docking with the highest affinity between the different search areas, is the site that is most likely to be that actual binding site. While this seems to be the true and obvious conclusion, it is not necessarily so. Due to the difference between the sites, there are some sites that have a larger inherent ability to have higher affinity. This is the case for the gridbox 8 search area as it is located in a cleft which will have larger possibility of bonds and interactions than the other search areas [41, 42]. While this of course makes the search area a good binding site, with its large possibility of bonds and interactions, it also means one must expect more from the site. The higher binding affinity is not necessarily better as one must expect more from it. Therefore it can not be concluded that the docking from search area 8 is the best just from the higher binding affinity.

Considering the theoretical impact of the different bindings, do they differ much, and are all likely? The different bindings bring with them different impacts, which can be theorized. The binding in domain IV makes one assume that a binding of EGF/TGF-α to Her2, will have similar effects to the binding of Herceptin to Her2. While a binding in the cleft between domain I and III would logically have a similar impact to the changes that happens when Her1 binds a ligand. But this is of course not likely to be the case, as the changes that the binding of ligands induces on Her1 can not take place in Her2. There is no tether to be broken between domain II and IV in Her2, and domain II does not need to be straightened. Further the binding of the ligands to the dimerization arm (Gridbox 9) would have an impact that contradicts the impact of the ligands in Her1 (the binding would hinder dimerization in Her2, but promote it for Her1). Lastly the binding to the area where IGG fab bind, there aren't any direct impacts. It would not block dimerization, but would not seem to induce it either as it doesn't interact with the dimerization arm. Overall the binding that seems to have largest potential on affecting Her2 is the one that binds in domain IV. The binding of EGF/TGF-α could have an effect similar to that of Herceptin, or somehow be involved in a binding protecting domain IV in a similar manner to the tether between domain II and IV in Her1.

Despite the fact that there are no concrete data from the docking experiments to determine which binding is the correct one , the binding in domain IV is the one that appears most likely to exist.

Which of the two ligands binds better to Her2?

Overall the binding affinities show that the bindings between TGF-α and Her2 are stronger than the bindings between EGF and Her2. This does not indicate that the binding of TGF-α is more likely than binding of EGF. There are many more factors to take into account, such as their availability in the extracellular space. Due to these extra factors, that are outside the capabilities of the docking programs considerations, there will be not made any assumptions which of the two ligands that will bind Her2 better. The likelihood of binding based on the binding affinity will therefore be treated as equal due to this uncertainty. Based on the information available about the docked structures, there is no reason to believe that either of the ligands would bind to Her2 better than the other.

Conclusion

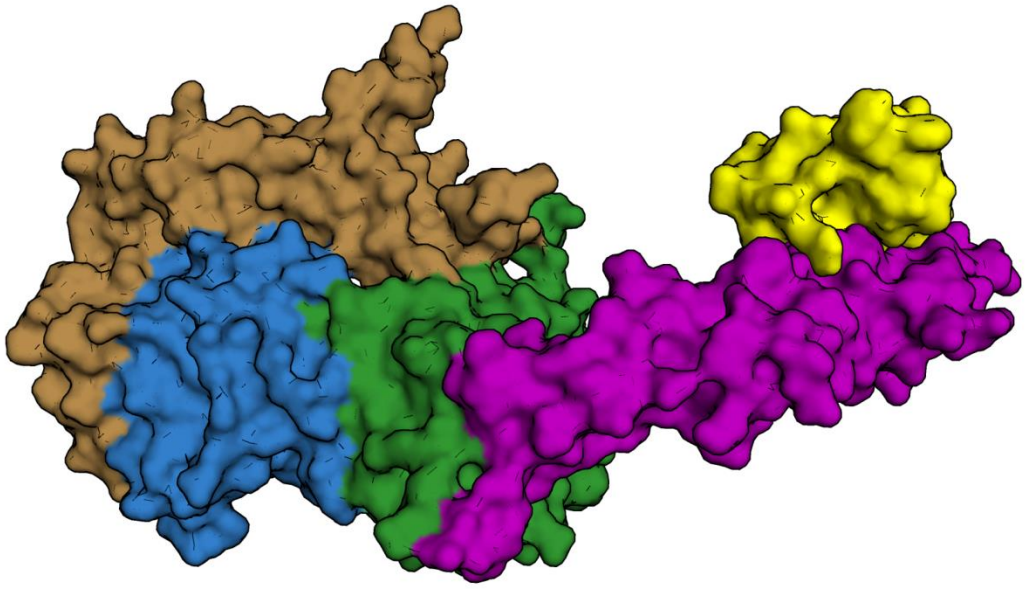

Figure 25: TGF-α (Yellow) Docked to Her2s domain IV (purple). Domain I of Her2 is colored blue, domain II brown and domain III green. The surface area is shown for both proteins. Image created in PyMOL [18]

In conclusion four bindings each for EGF and TGF-α to Her2 were found. The binding with highest binding affinity was in the cleft between domain I and III, similar to the binding that takes place between the two ligands and Her1. There was also found bindings for the two ligands to Her2 close to the binding sites of the two Her2-blocking cancer medicines Herceptin and Perjeta. These bindings are both strong bindings, with there being several theoretical impacts of these bindings that can later be looked into with lab research. The last binding of Her2 is in domain II where the IGG fab is bound in pdb structure 3WLW, this binding is weaker than the others and there are not any direct theoretical impact with the information available at this time. These bindings can be used as starting points for further research into Her2s relations with the ligands EGF and TGF-α.

Of all the docked structures found in the docking experiments, the one where the ligands were docked to domain IV of Her2 seems to be the best one overall (Shown in figure 25). This is because the binding affinity is high and there being several direct theoretical effects of the binding. Combined, these two attributes of this docked structure, makes it the one that seems the most likely. Even though the binding in domain IV is the more likely one, all of the bindings found in this thesis should be regarded as definitive possibilities for bindings between the two ligand proteins EGF and TGF-α and the oncoprotein Her2.

References:

1. Huang, S.-Y. (2014). Search strategies and evaluation in protein–protein docking: principles, advances and challenges. *Drug Discovery Today*, 19 (8): 1081-1096.
2. Bruno, O. V., Richard, E. & Maria, A. M. (2009). Structure-Based Virtual Ligand Screening: Recent Success Stories. *Combinatorial Chemistry & High Throughput Screening*, 12 (10): 1000-1016.
3. de Ruyck, J., Brysbaert, G., Blossey, R. & Lensink, M. F. (2016). Molecular docking as a popular tool in drug design, an in silico travel. *Advances and Applications in Bioinformatics and Chemistry : AABC*, 9: 1-11.
4. Bassaganya-Riera, J., Guri, A. J., Lu, P., Climent, M., Carbo, A., Sobral, B. W., Horne, W. T., Lewis, S. N., Bevan, D. R. & Hontecillas, R. (2011). Abscisic Acid Regulates Inflammation via Ligand-binding Domain-independent Activation of Peroxisome Proliferator-activated Receptor γ. *Journal of Biological Chemistry*, 286 (4): 2504-2516.
5. Venkatesan, K., Rual, J.-F., Vazquez, A., Stelzl, U., Lemmens, I., Hirozane-Kishikawa, T., Hao, T., Zenkner, M., Xin, X., Goh, K., II, et al. (2009). An empirical framework for binary interactome mapping. *Nature methods*, 6 (1): 83-90.
6. Trott, O. & Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *Journal of computational chemistry*, 31 (2): 455-461.
7. Klapper, L. N., Glathe, S., Vaisman, N., Hynes, N. E., Andrews, G. C., Sela, M. & Yarden, Y. (1999). The ErbB-2/HER2 oncoprotein of human carcinomas may function solely as a shared coreceptor for multiple stroma-derived growth factors. *Proceedings of the National Academy of Sciences of the United States of America*, 96 (9): 4995-5000.
8. Citri, A. & Yarden, Y. (2006). EGF–ERBB signalling: towards the systems level. *Nat Rev Mol Cell Biol*, 7 (7): 505-516.
9. Scaltriti, M. & Baselga, J. (2006). The Epidermal Growth Factor Receptor Pathway: A Model for Targeted Therapy. *Clinical Cancer Research*, 12 (18): 5268-5272.
10. Alroy, I. & Yarden, Y. (1997). The ErbB signaling network in embryogenesis and oncogenesis: signal diversification through combinatorial ligand-receptor interactions. *FEBS Letters*, 410 (1): 83-86.
11. Arkhipov, A., Shan, Y., Kim, E. T., Dror, R. O. & Shaw, D. E. (2013). Her2 activation mechanism reflects evolutionary preservation of asymmetric ectodomain dimers in the human EGFR family. *eLife*, 2: e00708.
12. Hu, S., Sun, Y., Meng, Y., Wang, X., Yang, W., Fu, W., Guo, H., Qian, W., Hou, S., Li, B., et al. (2015). Molecular architecture of the ErbB2 extracellular domain homodimer. *Oncotarget*, 6 (3): 1695-1706.
13. Ferguson, K. M., Berger, M. B., Mendrola, J. M., Cho, H.-S., Leahy, D. J. & Lemmon, M. A. (2003). EGF Activates Its Receptor by Removing Interactions that Autoinhibit Ectodomain Dimerization. *Molecular Cell*, 11 (2): 507-517.
14. Cho, H.-S. & Leahy, D. J. (2002). Structure of the Extracellular Region of HER3 Reveals an Interdomain Tether. *Science*, 297 (5585): 1330-1333.
15. Burgess, A. W., Cho, H.-S., Eigenbrot, C., Ferguson, K. M., Garrett, T. P. J., Leahy, D. J., Lemmon, M. A., Sliwkowski, M. X., Ward, C. W. & Yokoyama, S. (2003). An Open-and-Shut Case? Recent Insights into the Activation of EGF/ErbB Receptors. *Molecular Cell*, 12 (3): 541-552.

16. Ogiso, H., Ishitani, R., Nureki, O., Fukai, S., Yamanaka, M., Kim, J.-H., Saito, K., Sakamoto, A., Inoue, M., Shirouzu, M., et al. (2002). Crystal Structure of the Complex of Human Epidermal Growth Factor and Receptor Extracellular Domains. *Cell*, 110 (6): 775-787.
17. Garrett, T. P. J., McKern, N. M., Lou, M., Elleman, T. C., Adams, T. E., Lovrecz, G. O., Zhu, H.-J., Walker, F., Frenkel, M. J., Hoyne, P. A., et al. (2002). Crystal Structure of a Truncated Epidermal Growth Factor Receptor Extracellular Domain Bound to Transforming Growth Factor α. *Cell*, 110 (6): 763-773.
18. Schröedinger, L. (2010). *The PyMOL Molecular Graphics System, Version 1.8*. 1.8 ed.
19. Ferguson, K. M. (2008). A structure-based view of Epidermal Growth Factor Receptor regulation. *Annual review of biophysics*, 37: 353-373.
20. Fuller, S. J., Sivarajah, K. & Sugden, P. H. (2008). ErbB receptors, their ligands, and the consequences of their activation and inhibition in the myocardium. *Journal of Molecular and Cellular Cardiology*, 44 (5): 831-854.
21. Pandiella, A. & Massagué, J. (1991). Multiple signals activate cleavage of the membrane transforming growth factor-alpha precursor. *Journal of Biological Chemistry*, 266 (9): 5769-5773.
22. Carpenter, G. & Cohen, S. (1990). Epidermal growth factor. *Journal of Biological Chemistry*, 265 (14): 7709-7712.
23. Harris, R. C., Chung, E. & Coffey, R. J. (2003). EGF receptor ligands. *Experimental Cell Research*, 284 (1): 2-13.
24. Garrett, T. P. J., McKern, N. M., Lou, M., Elleman, T. C., Adams, T. E., Lovrecz, G. O., Kofler, M., Jorissen, R. N., Nice, E. C., Burgess, A. W., et al. (2003). The Crystal Structure of a Truncated ErbB2 Ectodomain Reveals an Active Conformation, Poised to Interact with Other ErbB Receptors. *Molecular Cell*, 11 (2): 495-505.
25. Mitri, Z., Constantine, T. & O'Regan, R. (2012). The HER2 Receptor in Breast Cancer: Pathophysiology, Clinical Use, and New Advances in Therapy. *Chemotherapy Research and Practice*, 2012: 743193.
26. Burstein, H. J. (2005). The Distinctive Nature of HER2-Positive Breast Cancers. *New England Journal of Medicine*, 353 (16): 1652-1654.
27. Hudis , C. A. (2007). Trastuzumab — Mechanism of Action and Use in Clinical Practice. *New England Journal of Medicine*, 357 (1): 39-51.
28. Cho, H.-S., Mason, K., Ramyar, K. X., Stanley, A. M., Gabelli, S. B., Denney, D. W. & Leahy, D. J. (2003). Structure of the extracellular region of HER2 alone and in complex with the Herceptin Fab. *Nature*, 421 (6924): 756-760.
29. Franklin, M. C., Carey, K. D., Vajdos, F. F., Leahy, D. J., de Vos, A. M. & Sliwkowski, M. X. (2004). Insights into ErbB signaling from the structure of the ErbB2-pertuzumab complex. *Cancer Cell*, 5 (4): 317-328.
30. Lax, I., Johnson, A., Howk, R., Sap, J., Bellot, F., Winkler, M., Ullrich, A., Vennstrom, B., Schlessinger, J. & Givol, D. (1988). Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in mouse cells, and differential binding of EGF and transforming growth factor alpha. *Molecular and Cellular Biology*, 8 (5): 1970-1978.
31. Lesk, A. M. (2010). *Introduction to Protein Science*. New York: Oxford University Press. 455 pp.
32. *AutoDock Vina Manual*. Available at: http://vina.scripps.edu/manual.html#faq (accessed: 16.02.17).
33. Montelione, G. T., Wuthrich, K., Burgess, A. W., Nice, E. C., Wagner, G., Gibson, K. D. & Scheraga, H. A. (1992). Solution structure of murine epidermal growth factor

determined by NMR spectroscopy and refined by energy minimization with restraints. *Biochemistry*, 31 (1): 236-49.
34. Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S. & Olson, A. J. (2009). AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *Journal of computational chemistry*, 30 (16): 2785-2791.
35. *About Abel*. (2012). Available at: http://www.uio.no/english/services/it/research/hpc/abel/more/ (accessed: 08.05.17).
36. Holm, L. & Rosenström, P. (2010). Dali server: conservation mapping in 3D. *Nucleic Acids Research*, 38 (suppl 2): W545-W549.
37. *Protein Structure Verification:Introduction*. Available at: http://swift.cmbi.ru.nl/gv/pdbreport/checkhelp/intro.html (accessed: 01.03.17).
38. *Protein Evolution*. (2011). Available at: http://bioinfosu.okstate.edu/pe/pe14.html (accessed: 01.03.17).
39. Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography*, 66 (Pt 1): 12-21.
40. Harvey, T. S., Wilkinson, A. J., Tappin, M. J., Cooke, R. M. & Campbell, I. D. (1991). The solution structure of human transforming growth factor α. *European Journal of Biochemistry*, 198 (3): 555-562.
41. Laskowski, R. A., Luscombe, N. M., Swindells, M. B. & Thornton, J. M. (1996). Protein clefts in molecular recognition and function. *Protein Science : A Publication of the Protein Society*, 5 (12): 2438-2452.
42. Lewis, R. A. (1991). [8] Clefts and binding sites in protein receptors. In vol. Volume 202 *Methods in Enzymology*, pp. 126-156: Academic Press.
43. Stang, E. (2016). Personal communication (17.10.16).
44. Huitfeldt, H. & Stang, E. (2016). Personal Communication (12.10.16).
45. Chung, A., Cui, X., Audeh, W. & Giuliano, A. (2013). Current status of anti-her2 therapies: predicting and overcoming herceptin resistance. *Clinical breast cancer*, 13 (4): 223-232.
46. Molina, M. A., Codony-Servat, J., Albanell, J., Rojo, F., Arribas, J. & Baselga, J. (2001). Trastuzumab (Herceptin), a Humanized Anti-HER2 Receptor Monoclonal Antibody, Inhibits Basal and Activated HER2 Ectodomain Cleavage in Breast Cancer Cells. *Cancer Research*, 61 (12): 4744-4749.
47. Skjeldal, L. (2016). NMBU. Unpublished works.
48. Kanai, R., Haga, K., Akiba, T., Yamane, K. & Harata, K. (2004). Biochemical and Crystallographic Analyses of Maltohexaose-Producing Amylase from Alkalophilic Bacillus sp. 707. *Biochemistry*, 43 (44): 14047-14056.
49. Berthold, C. L., Toyota, C. G., Richards, N. G. J. & Lindqvist, Y. (2008). Reinvestigation of the Catalytic Mechanism of Formyl-CoA Transferase, a Class III CoA-transferase. *Journal of Biological Chemistry*, 283 (10): 6519-6529.
50. Feinstein, W. P. & Brylinski, M. (2015). Calculating an optimal box size for ligand docking and virtual screening against experimental and predicted binding pockets. *Journal of Cheminformatics*, 7 (1): 18.

APPENDIX 1

Results from the DALI structural alignment and the DALI sequence alignment output:

Full output from DALI server:

```
No:  Chain   Z    rmsd lali nres  %id PDB  Description
1:  mol2-A 33.3  5.0  499   511   44 PDB  MOLECULE: EPIDERMAL GROWTH FACTOR RECEPTOR;
2:  mol2-A 18.9  2.2  159   511   31 PDB  MOLECULE: EPIDERMAL GROWTH FACTOR RECEPTOR;
3:  mol2-A 17.7  4.5  169   511   23 PDB  MOLECULE: EPIDERMAL GROWTH FACTOR RECEPTOR;
4:  mol2-A  8.2  2.9  122   511   15 PDB  MOLECULE: EPIDERMAL GROWTH FACTOR RECEPTOR;
5:  mol2-A  6.3  3.8  123   511   16 PDB  MOLECULE: EPIDERMAL GROWTH FACTOR RECEPTOR;
6:  mol2-A  6.1  2.6   76   511   16 PDB  MOLECULE: EPIDERMAL GROWTH FACTOR RECEPTOR;
7:  mol2-A  6.0  4.0  117   511   13 PDB  MOLECULE: EPIDERMAL GROWTH FACTOR RECEPTOR;
8:  mol2-A  5.5  3.6  106   511   12 PDB  MOLECULE: EPIDERMAL GROWTH FACTOR RECEPTOR;
9:  mol2-A  3.9  1.8   63   511   14 PDB  MOLECULE: EPIDERMAL GROWTH FACTOR RECEPTOR;
10: mol2-A  3.9  2.0   60   511   13 PDB  MOLECULE: EPIDERMAL GROWTH FACTOR RECEPTOR;
11: mol2-A  2.5  2.7   51   511   20 PDB  MOLECULE: EPIDERMAL GROWTH FACTOR RECEPTOR;
```

Sequence alignment of output 1 from DALI:

Notation: three-state secondary structure definitions by DSSP (reduced to H=helix, E=sheet, L=coil) are shown above the amino acid sequence. Structurally equivalent residues are in uppercase, structurally non-equivalent residues (e.g. in loops) are in lowercase. Amino acid identities are marked by vertical bars.

No 1: Query=mol1A Her2 (3WLWa) Sbjct=mol2A Her1(1IVOa) Z-score=33.3

```
DSSP  --LLEEELLLLLLLLllllHHHHHHHHHLLLLEEELLEEEELLLLLLLLHHHHHLLEE
Query --TQVCTGTDMKLRLPaspetHLDMLRHLYQGCQVVQGNLELTYLPTNASLSFLQDIQEV   58
ident     || ||  ||         |   |       | || |||| ||   |  ||||  ||||
Sbjct eeKKVCQGTSNKLTQLgtfedHFLSLQRMFNNCEVVLGNLEITYVQRNYDLSFLKTIQEV   60
DSSP  llLLEELLLLLLLLLLllhhhHHHHHHHHHLLLLEELLLEEELLLLLLLLLHHHHHLLEE


DSSP  LLLEEEELLLLLLLLLLLLLEELLLLLLLLLEEEEEELLLLLlLLLLLLLLLLLLLEEL
Query QGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGDPlSPGGLRELQLRSLTEIL  118
ident  ||||||  | |   ||  | | ||    |  |||||| | |     || ||  | | |||
Sbjct AGYVLIALNTVERIPLENLQIIRGNMYYENSYALAVLSNYDA-NKTGLKELPMRNLQEIL  119
DSSP  LLLEEEELLLLLLLLHHHLLEELLLLLLLLEEEEEELLLLL-LLLLLLLLLLLLLLLLL


DSSP  LLEEEEELLLLLLLHHHLLHHHHLLllLLLL-LEEELLLLL-LLLLLLLLLLLLLLLLLL
Query KGGVLIQRNPQLCYQDTILWKDIFHknNQLA-LTLIDTNRS-RACHPCSPMCKGSRCWGE  176
ident  | |    || ||    | | ||             |       |  | | |    |||
Sbjct HGAVRFSNNPALCNVESIQWRDIVS--SDFLsNMSMDFQNHlGSCQKCDPSCPNGSCWGA  177
DSSP  LLEEEEELLLLLLLHHHLLLLLLLL--LLLLlLLLLLLLLLlLLLLLLLLLLLLLLLLL


DSSP  LHHHLLLLLLLLLL-LLLLLLLL-LLHHhLLLLLLLLLLLLLLLLLLLLLLLEEELLEEE
Query SSEDCQSLTRTVCA-GGCARCKG-PLPTdCCHEQCAAGCTGPKHSDCLACLHFNHSGICE  234
ident   | || ||   ||     || |      ||| |||||||||  |||| |  |     |
Sbjct GEENCQKLTKIICAqQCSGRCRGkSPSD-CCHNQCAAGCTGPRESDCLVCRKFRDEATCK  236
DSSP  LLLLLLLLLLLLLLlLLLLLLLLlLLLL-LLLLLEELLLLLLLLLLLLEELLEEELLEEE
```

```
DSSP  LLLLLLEEELLLLLLEEELLLLEEELLEEELLLLLLLLLLLLLLLLLLLLLEEELLLl
Query LHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNYLSTDVGSCTLVCPLHNQEVTAe  294
ident   || |  ||  |     |||| | ||| ||  || ||  || |||   |     |
Sbjct DTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNYVVTDHGSCVRACGADSYEMEE-  295
DSSP  LLLLLLEEEELLLLEEEELLLLLEEELLEEELLLLLLLEELLLLLEELLLLLLLLLLL-


DSSP  LLLLLEEELlLLLLL-LLLLLLLLLLLLLLLLLHHHHHHHLLLLEELLLEEELHHHHLLL
Query DGTQRCEKCsKPCAR-VCYGLGMEHLREVRAVTSANIQEFAGCKKIFGSLAFLPESFDGD  353
ident ||   | ||     | || | |             ||  |  |  | | |  ||  | ||
Sbjct DGVRKCKKC-EGPCRkVCNGIGIGEFKDSLSINATNIKHFKNCTSISGDLHILPVAFRGD  354
DSSP  LLLLLLLLL-LLLLLlLLLLLLLHHHLLLLLLLHHHLHHHLLLLEELLLEEELHHHHHLL


DSSP  HHHLLLLLLLLHHHHHHHLLEEEELEEELLLLLLLLLLLLLLLLEEEEELLLLLEEEE
Query PASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLPDLSVFQNLQVIRGRILHNGAYSLT  413
ident     | || |  |    |  |||| | | |||    ||  | ||  ||||    |  ||
Sbjct SFTHTPPLDPQELDILKTVKEITGFLLIQAWPENRTDLHAFENLEIIRGRTKQHGQFSLA  414
DSSP  LLLLLLLLLHHHHHHHHHLLEELLLEEELLLLLLLLLLHHHLLLLEELLLLLLLLEEEE


DSSP  EELLLLLLLLLLLLLEELLLEEEEELLLLLLLLLLLLHHHHLLLLLLLEEEELLLLLHHH
Query LQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTVPWDQLFRNPHQALLHTANRPEDEC  473
ident    | |  |||||| |   |   |  |  ||   |  |  ||    |      || |  |
Sbjct VVSLNITSLGLRSLKEISDGDVTISGNKNLCYANTINWKKLFGTSGQKTKIISNRGENSC  474
DSSP  EELLLLLLLLLLLLLLLLLLEEEEELLLLLLLHHHLLHHHHLLLLLLLEELLLLLLHHHL


DSSP  HHLLLLLLLLLHHHLLLLLLLLLLLLLLLEELLLLEEllllllllllleeellleeell
Query VGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQECVeecrvlqglpreyvnarhclpch  533
ident    |  || ||    |||| |  || |    || |||
Sbjct KATGQVCHALCSPEGCWGPEPRDCVSCRNVSRGRECV----------------------  511
DSSP  LLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLL----------------------


DSSP  lllllllllllllllllllll
Query pecqpqngsvtcfgpeadqcva  555
ident
Sbjct ----------------------  511
DSSP  ----------------------
```

Residues that are involved in the binding of TGF-α and EGF to domain I and III of Her2 are shown in the sequence alignment above. They are colored according to the scheme below.

AaBb residues that are involved only in binding of EGF
AaBb Residues that are involved only in binding of TGF-α
AaBb Residues that are involved of binding both

APPENDIX 2

Gridboxes used in the thesis

Gridbox 1

Ligand(s) used: 2TGF
Receptor used: 3WLW

center_x = 46.667
center_y = 97.250
center_z = -25.611

size_x = 15
size_y = 18
size_z = 16

Gridbox 2

Ligand(s) used: 2TGF
Receptor used: 3WLW

center_x = 19.889
center_y = 4.667
center_z = -15.861

size_x = 30
size_y = 18
size_z = 12

Gridbox 3

Ligand(s) used: 2TGF
Receptor used: 3WLW

center_x = 8.843
center_y = 12.69
center_z = -10.3

size_x = 9
size_y = 32
size_z = 20

Gridbox 4

Ligand(s) used: 2TGF
Receptor used: 3WLW

center_x = -0.889
center_y = 14.303
center_z = -41.691

size_x = 12
size_y = 16
size_z = 16

Gridbox 5

Ligand(s) used: 2TGF
Receptor used: 3WLW

center_x = 14.886
center_y = 1.779
center_z = -31.686

size_x = 19
size_y = 14
size_z = 17

Gridbox 6 v1

Ligand(s) used: 2TGF
Receptor used: 1N8Y

center_x = 48.896
center_y = 144.968
center_z = 133.769

size_x = 16
size_y = 22
size_z = 30

Gridbox 6 v2

Ligand(s) used: 2TGF and 1EGF
Receptor used: 1N8Y

center_x = 57.704
center_y = 102.014
center_z = 131.137

size_x = 34
size_y = 32
size_z = 38

Gridbox 7

Ligand(s) used: 2TGF and 1EGF
Receptor used: 1N8Y

center_x = -18.825
center_y = 89.152
center_z = 56.621

size_x = 42
size_y = 30
size_z = 40

Gridbox 8

Ligand(s) used: 2TGF and 1EGF
Receptor used: 3WLW

center_x = 5.52
center_y = 4.721
center_z = -22.664

size_x = 31
size_y = 40
size_z = 42

Gridbox 9

Ligand(s) used: 2TGF and 1EGF
Receptor used: 3WLW

center_x = 45.766
center_y = 101.153
center_z = -26.556

size_x = 42
size_y = 36
size_z = 42

Test Gridbox 1

Ligand(s) used: 2TGF
Receptor used: 1WP6

center_x = 43.612
center_y = -17.062
center_z = 24.968

size_x = 36
size_y = 36
size_z = 36

Test Gridbox 2

Ligand(s) used: 2TGF
Receptor used: 1WP6

center_x = 14.503
center_y = -28.078
center_z = 49.243

size_x = 36
size_y = 36
size_z = 36

APPENDIX 3

Full tables of results from AutoDock Vina docking experiments.

Table 1: Output from AutoDock Vina experiment with gridbox 2 as search area and TGF as ligand

#	Affinity	Distance from best mode	
-	-	**Rmsd l.b.**	**Rmsd u.b.**
1	433,4 kcal/mol	0	0
2	435,7 kcal/mol	1,974	5,165

Table 2: Output from AutoDock Vina experiment with gridbox 6 v1 as search area and TGF as ligand

#	Affinity	Distance from best mode	
-	-	**Rmsd l.b.**	**Rmsd u.b.**
1	277,8 kcal/mol	0	0

Table 3: Output from AutoDock Vina experiment with gridbox 6 v2 as search area and TGF as ligand

#	Binding affinity	Distance from best mode	
-	-	**Rmsd l.b.**	**Rmsd u.b.**
1	-3,0 kcal/mol	0,000	0,000
2	-2,9 kcal/mol	2,205	3,727
3	-2,9 kcal/mol	3,514	10,966
4	-2,9 kcal/mol	3,480	12,163
5	-2,9 kcal/mol	2,209	3,336
6	-2,8 kcal/mol	3,382	14,084
7	-2,8 kcal/mol	3,669	10,790
8	-2,8 kcal/mol	3,114	12,809
9	-2,8 kcal/mol	3,210	14,384

Table 4: Output from AutoDock Vina experiment with gridbox 6 v2 as search area and EGF as ligand

#	Binding affinity	Distance from best mode	
-	-	**Rmsd l.b.**	**Rmsd u.b.**
1	-2,2 kcal/mol	0,000	0,000
2	-2,1 kcal/mol	2,526	5,066
3	-2,1 kcal/mol	4,622	11,061
4	-2,1 kcal/mol	3,413	15,683
5	-2,1 kcal/mol	4,636	11,210
6	-2,1 kcal/mol	4,151	16,734
7	-2,1 kcal/mol	4,159	16,681
8	-2,0 kcal/mol	2,670	5,430
9	-2,0 kcal/mol	3,999	16,944

Table 5: Output from AutoDock Vina experiment with gridbox 7 as search area and TGF as ligand

#	**Binding affinity**	**Distance from best mode**	
-	-	**Rmsd l.b.**	**Rmsd u.b.**
1	-3,5 kcal/mol	0.000	0.000
2	-3,3 kcal/mol	1,818	2,852
3	-3,3 kcal/mol	4,795	11,623
4	-3,3 kcal/mol	3,445	16,169
5	-3,2 kcal/mol	4,295	10,423
6	-3,1 kcal/mol	1,890	3,798
7	-3,1 kcal/mol	3,845	11,354
8	-3,1 kcal/mol	3,280	9,613
9	-3,0 kcal/mol	4,524	17,942

Table 6: Output from AutoDock Vina experiment with gridbox 7 as search area and with EGF as ligand

#	**Binding affinity**	**Distance from best mode**	
-	-	**Rmsd l.b.**	**Rmsd u.b.**
1	-2,1 kcal/mol	0,000	0,000
2	-1,9 kcal/mol	4,470	14.857
3	-1,8 kcal/mol	3,790	14,465
4	-1,8 kcal/mol	1,563	2,303
5	-1,8 kcal/mol	4,842	16,297
6	-1,8 kcal/mol	4,472	15,227
7	-1,7 kcal/mol	4,598	11,681
8	-1,7 kcal/mol	4,302	15,248
9	-1,7 kcal/mol	4,366	11,520

Table 7: Output from AutoDock Vina experiment with gridbox 8 as search area and TGF-α as ligand

#	**Binding affinity**	**Distance from best mode**	
-	**-**	**Rmsd l.b.**	**Rmsd u.b.**
1	-4,9 kcal/mol	0.000	0,000
2	-4,7 kcal/mol	2,312	3,539
3	-4,6 kcal/mol	3,875	18,430
4	-4,5 kcal/mol	3,800	18,306
5	-4,5 kcal/mol	4,387	18,999
6	-4,5 kcal/mol	3,817	12,268
7	-4,4 kcal/mol	4,340	21,952
8	-4,4 kcal/mol	3,783	19,107
9	-4,3 kcal/mol	4,047	18,144

Table 8: Output from AutoDock Vina experiment with gridbox 8 as search area and EGF as ligand

#	**Binding affinity**	**Distance from best mode**	
-	-	**Rmsd l.b.**	**Rmsd u.b.**
1	-4,6 kcal/mol	0,000	0,000
2	-4,6 kcal/mol	1,417	2,057
3	-4,3 kcal/mol	5,245	17,289
4	-4,2 kcal/mol	2,143	6,095
5	-4,1 kcal/mol	4,069	20,021
6	-4,1 kcal/mol	3,940	20,472
7	-4,1 kcal/mol	4,643	16,001
8	-4,1 kcal/mol	3,578	6,302
9	-4,1 kcal/mol	4,597	16,027

Table 9: Output from AutoDock Vina experiment with gridbox 9 as search area and TGF-α as ligand

#	**Binding affinity**	**Distance from best mode**	
-	-	**Rmsd l.b.**	**Rmsd u.b.**
1	-3,7 kcal/mol	0,000	0,000
2	-3,7 kcal/mol	1,790	2,621
3	-3,4 kcal/mol	4,265	15,804
4	-3,4 kcal/mol	1,854	3,010
5	-3,4 kcal/mol	4,970	18,046
6	-3,4 kcal/mol	4,614	15,843
7	-3,3 kcal/mol	5,347	16,789
8	-3,3 kcal/mol	3,942	16,210
9	-3,3 kcal/mol	3,741	17,386

Table 10: Output from AutoDock Vina experiment with gridbox 9 as search area and EGF as ligand

#	**Binding affinity**	**Distance from best mode**	
-	-	**Rmsd l.b.**	**Rmsd u.b.**
1	-3,2 kcal/mol	0,000	0,000
2	-3,1 kcal/mol	2,232	4,292
3	-3,1 kcal/mol	2,538	4,140
4	-3,1 kcal/mol	2,874	5,693
5	-3,1 kcal/mol	3,105	6,682
6	-3,0 kcal/mol	4,620	14,519
7	-2,9 kcal/mol	2,541	4,344
8	-2,9 kcal/mol	2,903	6,220
9	-2,9 kcal/mol	3,917	18,922

Table 11: Output from AutoDock Vina experiment with test gridbox 1 as search area and TGF-α as ligand

#	**Binding affinity**	**Distance from best mode**	
-	-	**Rmsd l.b.**	**Rmsd u.b.**
1	-0,9 kcal/mol	0,000	0,000
2	-0,8 kcal/mol	1,737	2,351
3	-0,6 kcal/mol	1,971	3,134
4	-0,6 kcal/mol	1,857	2,634
5	-0,4 kcal/mol	1,723	2,337
6	0,5 kcal/mol	2,545	3,687
7	0,5 kcal/mol	1,768	2,555
8	1,6 kcal/mol	3,162	11,231
9	1,6 kcal/mol	2,984	11,098

Table 12 Output from AutoDock Vina experiment with test gridbox 2 as search area and TGF-α as ligand

#	**Binding affinity**	**Distance from best mode**	
-	-	**Rmsd l.b.**	**Rmsd u.b.**
1	-4,5 kcal/mol	0,000	0,000
2	-4,3 kcal/mol	4,313	18,808
3	-4,3 kcal/mol	4,855	18,207
4	-4,2 kcal/mol	5,603	15,742
5	-4,1 kcal/mol	1,856	2,816
6	-4,1 kcal/mol	4,397	17,434
7	-4,1 kcal/mol	6,081	16,407
8	-4,1 kcal/mol	4,443	19,226
9	-4,1 kcal/mol	4,097	11,141

Norges miljø- og biovitenskapelig universitet
Noregs miljø- og biovitskapelege universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway