FH CAMPUS WIEN
UNIVERSITY OF APPLIED SCIENCES

# Prediction Tools for Protein-Protein Binding

Machine Learning Tools for mRNA function prediction

**Bachelor Thesis**

Submitted in partial fulfillment of the requirements for the degree of:
**Bachelor of Science in Engineering**

to the University of Applied Sciences FH Campus Wien
Bachelor Degree Program: Computer Science and Digital Communications

**Author:**
Jakob Schörg

**Student Identification Number:**
1810475023

**Supervisor:**
DI Dr. techn. Mugdim Bublin

**Date:**
17. 03. 2021

Declaration of authorship:

I declare that this Bachelor Thesis has been written by myself. I have not used any other than the listed sources, nor have I received any unauthorized help.

I hereby certify that I have not submitted this Bachelor Thesis in any form (to a reviewer for assessment) either in Austria or abroad.

Furthermore, I assure that the (printed and electronic) copies I have submitted are identical.

Date: 17.03.2021..... Signature: Jakob Schörg

# Abstract

Antibody-antigen complexes represent a central role in drug researching. The study of proteins allows deeper insight in protein binding mechanisms. However, experimental protein-protein docking analysation is time consuming and expensive. Development in computational antibody-antigen structure prediction and increasing processing power paved the way for drug researching. In this work the efficiency of recent binding prediction tools is stated by examining the linked structure of several selected antibodies and antigens. Thereafter, the results are compared to each other and a resume of these tools is made. For this investigation the prediction tools pyDockWeb and PatchDock were used on the immunoglobins IgE PA12P3F10 and PA13P1H08 linked to each of the allergens Arah1, Arah 2, Arah 3.

# List of Abbreviations

| | |
|---|---|
| Ab | Antibody |
| Ag | Antigen |
| α-helix | Alpha Helix |
| β-helix | Beta Helix |
| TM | Template-based modelling |
| FM | Free modelling |
| ML | Machine Learning |
| FASTA | Fast Alignment |

# Key Terms

Antibody-antigen binding

Immunoglobins

Antigens

Protein structure prediction

Machine Leaning

AlphaFold

DNA

pyDockWeb

PatchDock

# Contents

# 1. Introduction

Our body is constantly exposed to a harsh environment hosting a various assortment of destructive allergens. The resident immune system is steadily on alert and intervenes in case of detecting a foreign body. If an alien particle circulates in the system, an antibody, perfectly tailored to this intruder is sent and attaches to it. The binding with the antigen turns it innocuously and then can simply be removed from the body.

Besides some exceptions, almost all antibodies and antigens are proteins. Proteins are the substance building material in all living systems. Their variable domains provide a wide range of different functions and therefore a high variety of applications. Proteins are the product of the decoding process of cellular DNA information. Each gene includes an individual protein structure code in cellular DNA. Every protein consists of a varying number of amino acids arranged in a chain. Multiple amino acids are bound within a protein, forming a long chain by peptide connections. Peptide bonds form a biochemical reaction that removes a molecule of water as it unites an amino acid group with a neighboring amino acid carboxyl group.

The order of an amino acid sequence resolves the folding process in total on four different levels. The primary protein structure is called the linear sequence of amino acids within a protein Fig. 1 (a). Only 20 amino acids, each having a special side chain, are used in the construction of proteins. There are various chemistries in the side chains of amino acids. The chemical characteristics of amino acid side chains are important for protein structure, since they can bind to each other in a certain shape or conformation, to maintain a protein length.

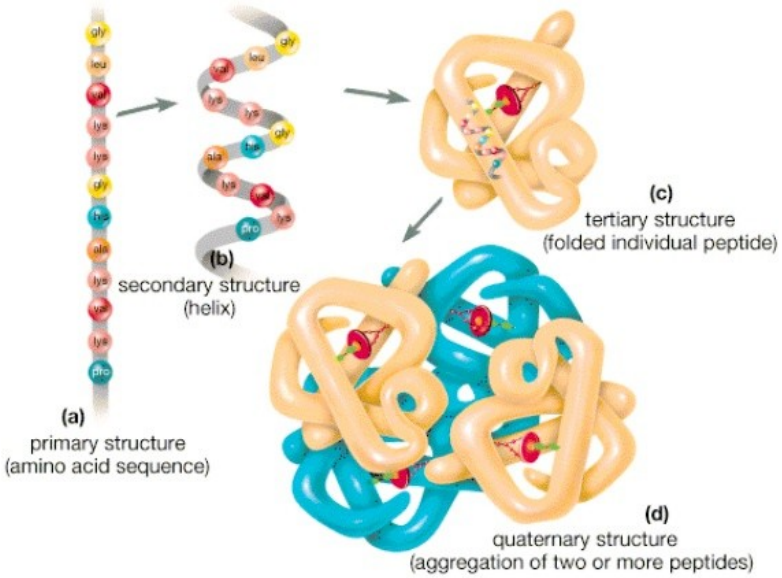



Fig. 1: The 4 levels of protein folding

The sequence and placement of amino acids in such protein guides in which bends and folds occur in the protein due to side chain interactions. Hydrogen binding in adjacent regions of the protein chain between amino groups and carboxy groups often leads to such folding patterns identified as alpha helices and beta sheets Fig. 1 (b). In a linear chain of amino acids, the ensemble of formations and folds form the 3-dimensional structure of a protein (c). The last folding step applies macromolecules that have many polypeptide chains or subunit systems to one protein (d).

The variable V sections of any single molecule of an antibody are distinct from the others. However, sequence heterogeneity is not uniformly spread across the V zones, but concentrates on some V-area segments. In both the VH and VL domains, three segments of unique heterogeneity can be defined. Antibodies have a biochemical role to bind to and facilitate pathogens. A small area on the surface of a large molecule, like a protein, is typically recognised by an antibody. The antibody structure is referred to as an epitope. Antibodies raised against peptides are often observed to bind to the normal folded protein of a protein that corresponds to part of his sequence. In certain cases synthetic peptides may be used in vaccines intended to increase antibodies to a pathogen protein. [ CITATION Jan01 \l 1031 ]

However, creating synthetic antibodies experimentally demands for high resource costs. Therefore, machine learning algorithms were introduced to find a customized antibody for the antigen. Progress of computation power and development of algorithms enhanced protein-protein binding prediction tools drastically. In this work two tools for binding structure prediction were selected and tested on a set of proteins.

.

# 2. Protein Fundamentals

Prior to Antibody-Antigen Binding is the understanding of fundamental keywords and processes in biology to be able to comprehend the results. The In this section a basic introduction to proteins, their structure, their coherent function and further antibodies and antigens is provided.

## 1..1. Proteins

A Protein consists of macromolecules of one or more amino acid chain residues that varies in length, structure and function. They are the most prevalent part of living systems. Their various functions are categorized in structural, regulatory, contractile, defensive and they provide capabilities for transport, storage and membranes. A protein is a polymer, built up by a chain of small organic molecules/monomers, called amino acids. [ CITATION Conas \l 1031 ]

### 2nd.1st.1st Amino Acids

The fundamental unit of a protein is an amino acid. About 700 amino acids are found in nature.

Their affiliated basis is a central carbon atom tied to an amino group, a carboxyl group, and a hydrogen atom. Additionally, each amino acid owns a single atom, or a group of atoms (R group) bound to the central carbon atom.

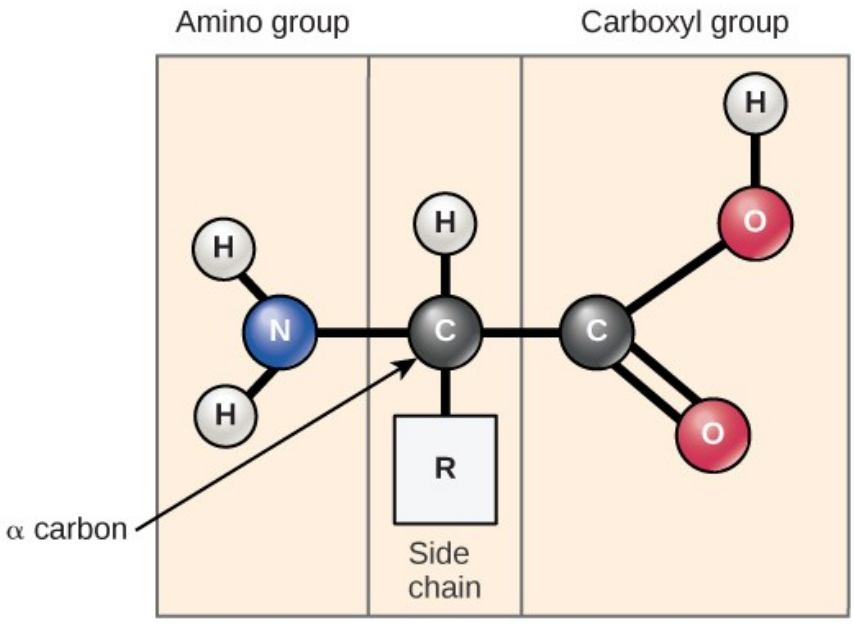



Fig. 2: Visualisation of an amino acid on atomic level

The R group, also known as side chain, rules the behavior of the monomer. There are four formations of amino acids: acidic, basic, polar, nonpolar.

The representation of an amino acid is a single uppercase or three letter abbreviation.

## 2..1..2. Peptides

As already stated, consists a protein of an amino acid sequence**.** In order to form an interrelated chain, amino acids must connect to each other.

The bond between two acids is called a peptide bond. It is caused by an amino acids carboxyl group and the other amino group bind together. A water molecule is released during this process.

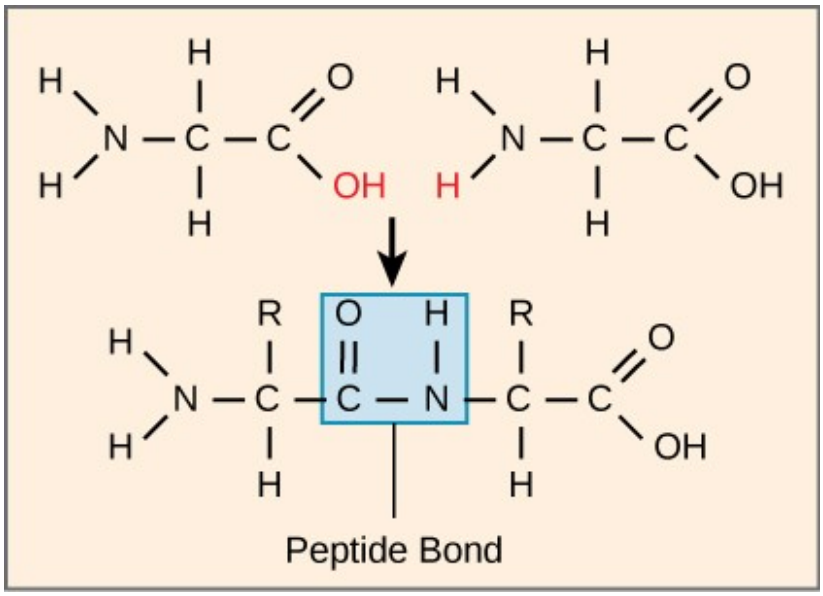



Fig. 3: Visualization of the peptide bond.

The result of two bound amino acids is a peptide and more than two is a polypeptide. A peptide is characterised by a free group of aminos. On one end is it termed N or amino terminal and on the other end a free carboxyl group called C or carboxyl terminal. Thus, a protein is a linking of peptides or polypeptides.

## 2..1..3.Types and Functions

In the table below the proteins are categorized in 7 different types with their coherent function.

| Type | Examples | Function |
|---|---|---|
| Digestive Enzymes | Amylase, lipase, pepsin, trypsin | Help to digest nutrients into monomeric units by catabolizing food |
| Transport | Hemoglobin, albumin | Transport substances in the blood or lymph across the whole body |

| | | |
|---|---|---|
| Structural | Actin, tubulin, keratin | Build various structures, such as cytoskeleton |
| Hormones | Insulin, thyroxine | Arranges the body's activity coordination |
| Defense | Immunoglobulins | Defends the body from alien diseases |
| Contractile | Actin, myosin | For contraction of the muscles |
| Storage | Legume storage proteins, egg white (albumin) | Feeds the embryo and the seedling early in its development |

Tab. 1: Proteins Types and Functions

## 2..1..4. Protein Structure

The protein structure affects its function directly. The result of the folding process defines the protein structure, and it takes part on four levels (primary, secondary, tertiary, quaternary). Proteins also frequently get assistance from protein helpers (chaperones) in the folding operation which disjoins after folding.

### Primary Structure

The primary structure establishes the order of the amino acids in a chain. The individual divisions are defined by the encoding gene. If a building molecule of the gene is changed a different (distinct**)** amino acid is attached to the polypeptide, which induces a modification of the protein structure and function.

### Secondary Structure

The secondary structure specifies **(indicates)** the local folding. The prevailing shapes are the α-helix and β-pleated sheet, sticked together by hydrogen bonds.

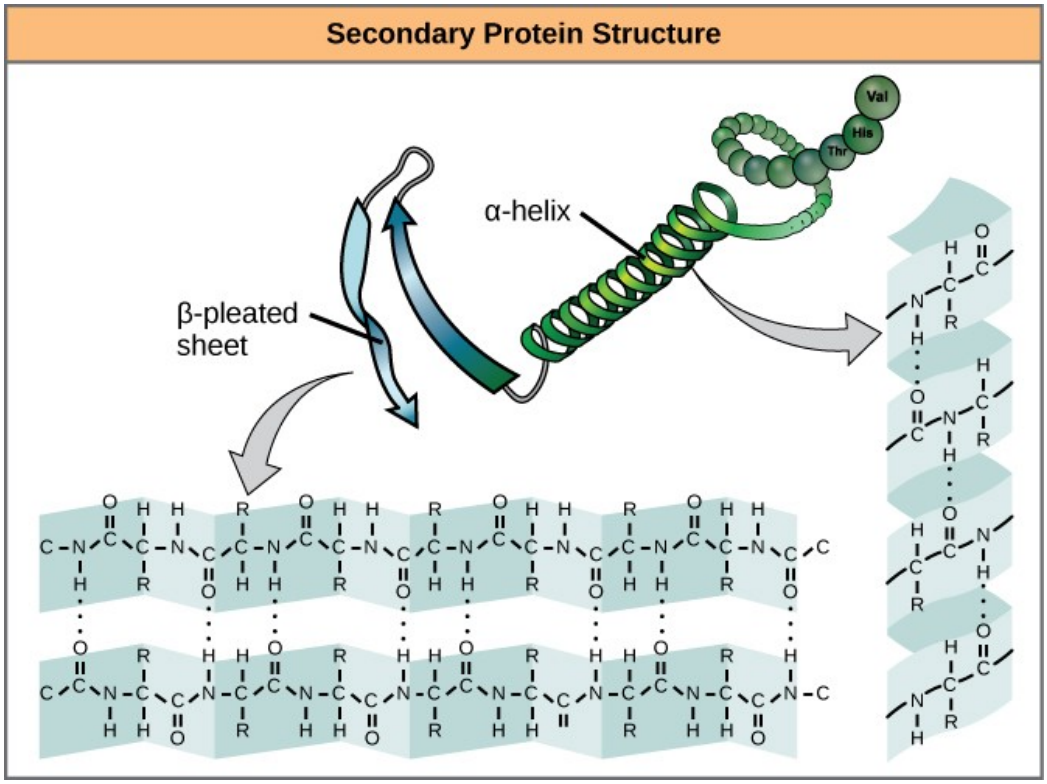



Fig 4. Secondary Protein Structure

An α-helix keeps 3,6 amino acid residues in each helical turn and the R groups bulge out of the chain.

The β-pleated sheet is compound by hydrogen bonds sandwiched by atoms on the backbone of the polypeptide chain.

### Tertiary Structure

The fundamental cause for this process are the connecting R groups which can even counteract the outcome of the hydrogen bonds on the secondary level. Besides, chemical procedures take place on the polypeptide chain.

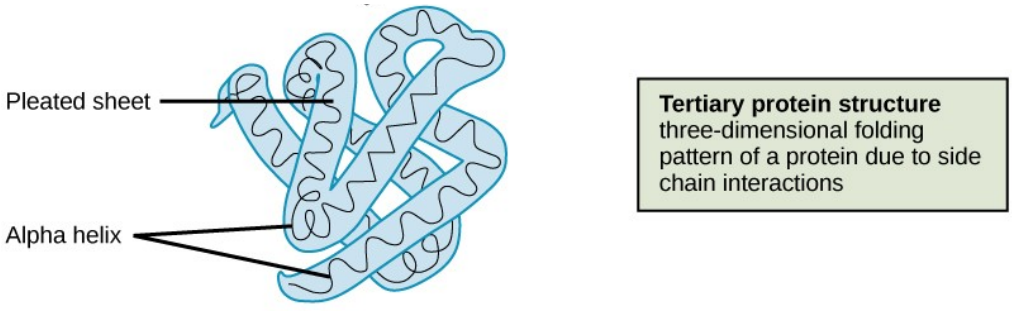



Fig 5. Tertiary Protein Structure

### Quaternary Structure

Proteins by nature consist of manifold weakly linked polyepitopes, which form the quaternary structure. In the illustration below the four levels of protein structure folding can be observed.

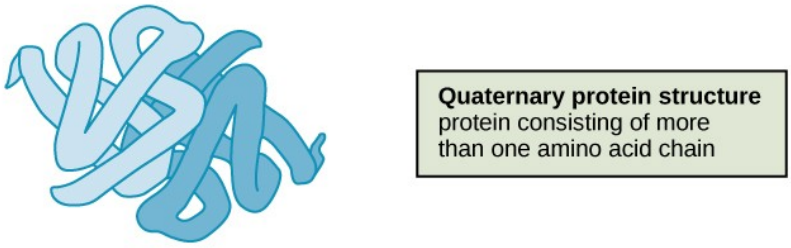



Fig 6. Quaternary protein structure

The final form of a newly synthesized protein is usually the most energy-efficient. As proteins fold, they measure a range of conformations, which is certain and compact, until they are done. Thousands of noncovalent bonds between amino acids are sustained by folding proteins. Chemical forces between the protein and its immediate environment often contribute to the form and stability of the protein. [ CITATION Sci20 \l 1031 ]

### Denaturation

The denaturation, known as the course of action when a protein is exposed to environmental changes evoking a transformation in structure without alternation of its primary sequence. It is mostly reversable when abolishing the initiator.

# 2..2. Antibodies & Antigens

## 2..2..1.Antibody

Antibodies otherwise known as immunoglobins are made up of four peptides arranged in a Y-shape and are produced by white blood cells. It contains a paratope, which is the connective link to an epitope, the antigens linkage. The binding of the paratope with the epitope countervails antigens or at least tags the harmful protein. Subsequently the tagged protein will be neutralized by the immune system. [ CITATION Ann17 \l 1031 ] [CITATION Kat20 \l 1031 ]

In the Figure below the structure of an antibody is demonstrated. An antibody consists of two identical light chains (L) and two identical heavy chains (H), kept together by disulfide bonds. A constant region (C) and a variable region (V), where the antigen-binding occurs, are included in each chain. The constant region classifies the antibody, whereas five classes exist (IgM, IgG, IgA, IgE, IgD). [ CITATION Bri20 \l 1031 ]

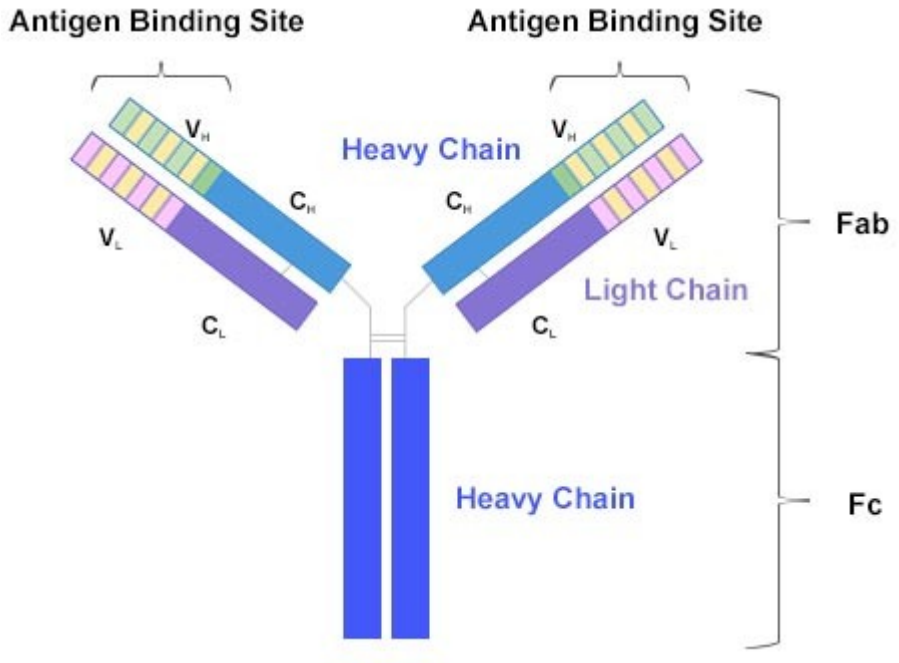



Fig 7. Antigen Binding Site

In general, there exist three different methods antibodies make use of to fight off antigens:

**Neutralization**

Antibody binding is used to neutralise foreign particles to prevent cell interaction resulting in cell damage and invasion.

**Opsonization**

Antibodies can serve as tags recruiting immune cell mediators to phagocytose the elimination of foreign substances.

**Complement**

Antibodies may coat foreign particles and activate the complementary mechanism that directly lyses or further induces opsonisation of alien materia.

## 2..2..2.Antigen

Antigens are more often than not proteins, originated either within the body (autoantigens) or externally, harming the living system. Its intrusion in a living system triggers an immune response. The immune system hast to bind an antibody to this antigen accordingly. As already stated, is the binding site of the antigen called epitope. [ CITATION Ann17 \l 1031 ]

Particular antigens possess a resembling constant bottom and stem structure like an antibody. The binding sites are variable regions and strongly differ. [ CITATION Bri20 \l 1031 ]

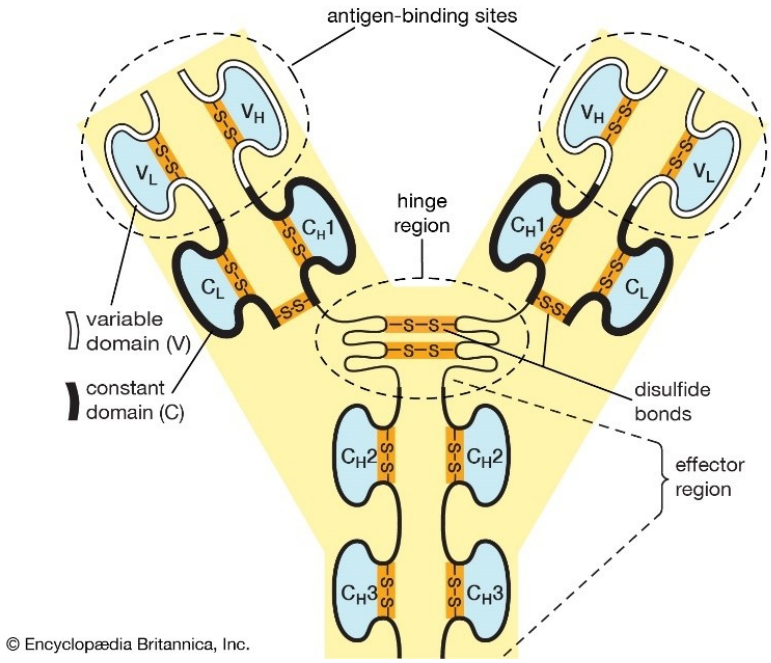



Fig. 8 Structure of Antigen

# 2..3. Antibody-Antigen Binding

Antibodies are large proteins that function by binding themselves to an antigen, which is a pathogen's specific location. The immune system is mobilized to target and neutralize the pathogen as a result of the binding. Antibodies have a special structure that grants them to be highly selective while still being low immunogenic.

The immune response is activated by the fragment crystallizable region (Fc region), which is species specific**,** i.e. human Fc region does not elicit an immune response in humans. To be able to bind to antigens of different types, the fragment antigen-binding region (Fab region) must be highly variable.[ CITATION HAD21 \l 1031 ]

The paratope, the small part of the antibody's Fab region, is made up of six remarkably versatile loops called complementarity-determining regions (CDRs) or hypervariable loops. These can change their sequence and conformation to bind to different antigens.

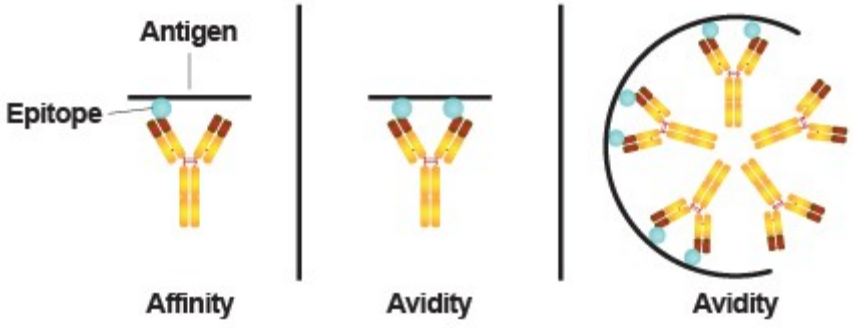



Fig 9 Antibody Binding Variations

The links that carry the antigen at the combined site of any anticorps are non-covalent and are thus naturally reversible.
These relations can consist of hydrogen links, electrostatic bonds or the forces of the Van der Waals.
Multiple bond formation is typically observed, ensuring that the binding between antibody and antigen is relatively strong.
There are very few portions of the molecules involved, usually just a few aminos, for the unique binding between the determinant of a cell and the antigen combination site.
In antigen-antibody reactions, the locations are important because a special bond must resolve the repulsion between the two molecules.
When the epitope encounters the paratope, ionic and hydrophobic forces attract each other.
They help resolve their hydration energies and permit the expulsion of water molecules as paratope and epitope approach each other. Van der Waal forces even strengthen this seduction.

# 3. Computational docking

As determining a protein structure experimentally is very expensive and time-consuming and computational power has risen for a lower cost, machine learning algorithms have become more widespread**.** The large difference between the amount of sequences available and protein structures which have been experimentally solved may be overcome by prediction models.

A major difficulty of structural genomics is to obtain structural knowledge from sequence alone. The unit of covariation in most algorithms is in point of fact pairs of single columns. Standard deviation between pairs of columns has been used to identify points of inter-protein docking and analyse for packing unique to α-helixes to α-helixes distances, as well as to find errors in alignments.

Other approaches also attempt to extend knowledge from the SCA (statistical coupling analysis) algorithms to several sets of columns in one example of this approach. Prediction of theoretical binding structure can be split in two specific camps: free modeling (ab-initio) and homology (template based) modelling.[ CITATION Kyl20 \l 1031 ]

## 3..1. Template based modeling

Template based modeling remains as the most accurate method for computational protein binding prediction. Based on experimental data the algorithm determines the protein structure of similar undiscovered proteins. The difficulties lie in choosing the ideal template protein and the efficiency of the modeling algorithm. [ CITATION LiC17 \l 1031 ]

## 3..2. Free Modeling

The most difficult (**demanding)** challenge in 3-dimensional structure prediction is to model proteins with little or minimal resemblance to actual structures. Ab-initio modeling predicts folds from physical chemistry principles.

The first model of a reasonable precision was probably established in CASP4. Henceforth, CASP's ab initio prognosis has seen incessant progress, but mainly for small proteins (120 residues or under). In CASP11 for the first time**,** the new fold protein was developed with an before unpresented precision for targets of this size (256 residue, sequence identity for known structures ·5 percent). In addition, the experiments of CASP11 and CASP12 (2014, 2016) showed a new tendency to build improved models using the forecasted contacts.

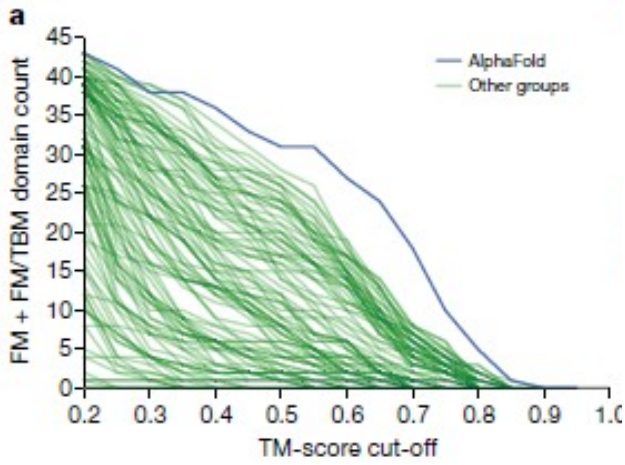
CASP13 saw a further significant increase in the precision of template free models due primarily to advanced deep learning artificial intelligence techniques and the estimation of residue distances at various thresholds. More than 20% improvement in accuracy was seen in the best models (including AlphaFold) submitted on the Free Modeling Goals. [ CITATION Pro20 \l 1031 ]

### 3..2..1.AlphaFold

AlphaFold is a free modelling machine learning algorithm tool. It mainly predicts the distances between different residue pairs. In contrast to the predictions of touch, this technique provides additional structural details. This knowledge enables us to try to create a capacity to explain the shape of a protein.

A simple gradient decent algorithm can further optimise it, which produces structures without complex sampling procedures. All of this is a highly precise AlphaFold scheme, even for less homologous sequences.

In the CASP13 (Critical Assessment of Protein Structure Prediction) assessment, Considering the fact that AlphaFold competed also in the template based category, it could predict a higher level of FM domains accuracy than any other group (ttt c), especially if we consider the range of TM 0.6 to 0.7 shown in the graphic below (ttt a). This range demonstrates the degree of correspondence between the given structure for computation to the natural structure.



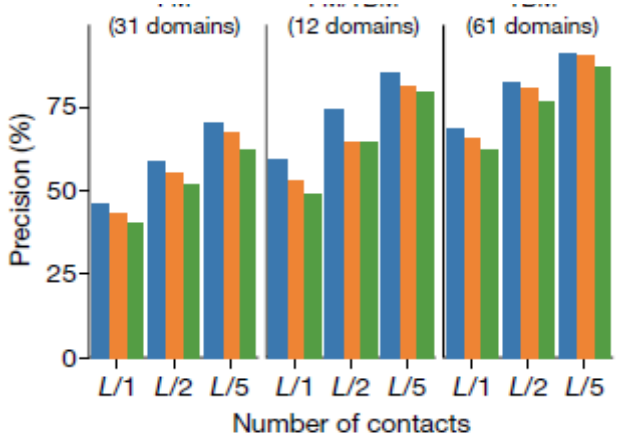



Fig 10. CASP 13

## 3..3. Formats

In this work I came across two important file formats:

1. FASTA: An amino acid chain file format with a short sequence description.
2. PDB: This file format is used in the PDB (protein data base) database and includes coordinates of proteins on an atomic level.

### 3..3..1. FASTA format

The FASTA format is a text-based file format to display exhibit nucleotide or peptide segments, in which the base pairs or amino acids are displayed by means of single-letter codes. A single-line summary followed by sequence data lines in the FASTA format starts. A greater symbol (">") is used in the first column to differentiate the definition line from the sequence data. The length of all text lines is suggested to be less than 80 characters. A typical IUB/IUPAC amino acid and nucleic acid codes are required to represent sequences.

### 3..3..2. PDB format

A PDB file has determined biological macromolecular structures in three dimensions. The data comprises atomic co-ordinates, factors of crystallographic structure and experimental NMR data. Apart from codes, the molecules names, primary and secondary structure details are also included in each deposition.

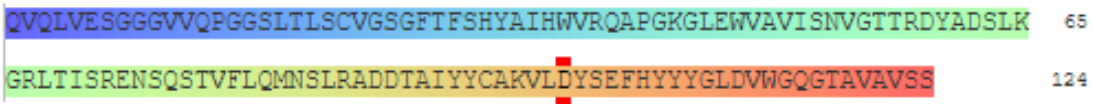
In several lines each PDB file is shown. There are 80 columns for each line in the PDB entry file. An end-of-line indicator should be the last character of each PDB input. Each line begins with 6 digits for the record name, which are segregated with a blank and can be on its own.

# 4. Results

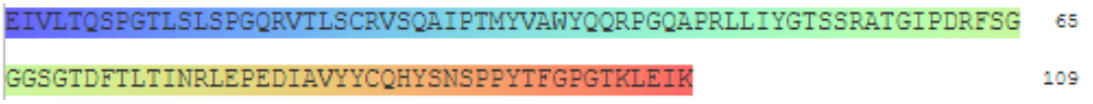
The target of this work is to compute the protein-protein docking structure out of antibodies and antigens with different tools (pyDockWeb, patchDock).

The proteins used in this example were the antibodies PA12P3F10, PA13P1H08 and the antigens Ara h 1 (sp|P43238|ALL12_ARAHY Allergen), Ara h 2 (sp|Q6PSU2|CONG7_ARAHY Conglutin-7) and Ara h 3 (tr|Q8LKN1|Q8LKN1_ARAHY Allergen). Their amino sequences are listed below:

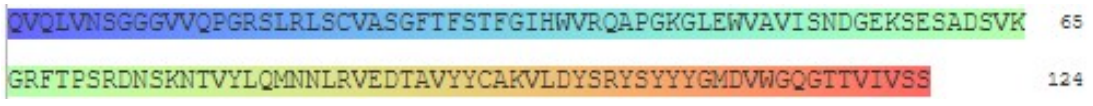
- PA12P3F10:
    - Heavy Chain variable region



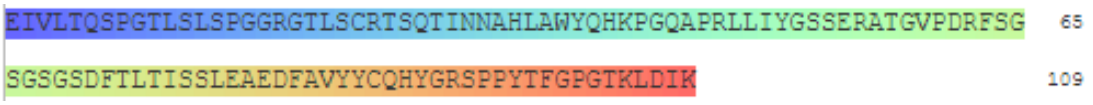

    - Light Chain variable region




- PA13P1H08:
    - Heavy Chain variable region




    - Light Chain variable region




- Ara h 1:

Description line: >sp|P43238|ALL12_ARAHY Allergen

Ara h 1, clone P41B OS= Arachis hypogaea

```
QRSPPGERTRGRQPGDYDDDRRQPRREEGGRWGPAGPRERERREEDWRQPREDWRRPSHQQPRKIR  65
PEGREGEQEWGTPGSHVREETSRNNPFYFPSRRFSTRYGNQNGRIRVLQRFDQRSRQFQNLQNHR 130
IVQIEAKPNTLVLPKHADADNILVIQQGQATVTVANGNNRKSFNLDEGHALRIPSGFISYILNRH 195
DNQNLRVAKISMPVNTPGQFEDFFPASSRDQSSYLQGFSRNTLEAAFNAEFNEIRRVLLEENAGG 260
EQEERGQRRWSTRSSENNEGVIVKVSKEHVEELTKHAKSVSKKGSEEEGDITNPINLREGEPDLS 325
NNFGKLFEVKPDKKNPQLQDLDMMLTCVEIKEGALMLPHFNSKAMVIVVVNKGTGNLELVAVRKE 390
QQQRGRREEEEDEDEEEEGSNREVRRYTARLKEGDVFIMPAAHPVAINASSELHLLGFGINAENN 455
HRIFLAGDKDNVIDQIEKQAKDLAFPGSGEQVEKLIKNQKESHFVSARPQSQSQSPSSPEKESPE 520
KEDQEEENQGGKGPLLSILKAFN                                            543
```

- Ara h 2:

Description line: >sp|Q6PSU2|CONG7_ARAHY Conglutin

7 Ara h 2.0201) OS= Arachis hypogaea

```
RQQWELQGDRRCQSQLERANLRPCEQHLMQKIQRDEDSYGRDPYSPSQDPYSPSQDPDRRDPYSP  65
SPYDRRGAGSSQHQERCCNELNEFENNQRCMCEALQQIMENQSDRLQGRQQEQQFKRELRNLPQQ 130
CGLRAPQRCDLEVESGGRDRY                                              151
```

- Ara h 3:

Description line: >tr|Q8LKN1|Q8LKN1_ARAHY Allergen

Ara h 3 /Arah4 Arachis hypogaea OX=3818 PE=3 SV=1

```
ISFRQQPEENACQFQRLNAQRPDNRIESEGGYIETWNPNNQEFECAGVALSRLVLRRNALRRPFY  65
SNAPQEIFIQQGRGYFGLIFPGCPSTYEEPAQQGRRHQSQRPPRRFQGQDQSQQQQDSHQKVHRF 130
DEGDLIAVPTGVAFWMYNDHDTDVVAVSLTDTNNNDNQLDQFPRRFNLAGNHEQEFLRYQQQSRR 195
RSLPYSPYSPQTQPKQEDREFSPRGQHGRRERAGQEQENEGGNIFSGFTPEFLAQAFQVDDRQIL 260
QNLRGENESDEQGAIVTVRGGLRILSPDRKRRQQYERPDEEEEYDEDEYEYDEEERQQDRRRGRG 325
SRGSGNGIEETICTASFKKNIGRNRSPDIYNPQAGSLKTANELQLNLLILRWLGLSAEYGNLYRN 390
ALFVPHYNTNAHSIIYALRGRAHVQVVDSNGDRVFDEELQEGHVLVVPQNFAVAGKSQSENFEYV 455
AFKTDSRPSIANLAGENSFIDNLPEEVVANSYGLPREQARQLKNNNPFKFFVPPSEQSLRAVA   518
```

**The** protein sequences from above first had to be converted into PDB file format (protein data bank file format), therefore the Swiss Model was used in this work.

SWISS-Model is a tool for protein structure homology modeling providing a web service interface. For the input a simple amino acid chain is required in the FASTA format. The output can be downloaded as zip-file and includes among others a PDB-file and a visual 3D model of the protein structure in PNG-format.

Next all variants of antibody-antigen binding (with the specified proteins above) were went through with both tools pyDockWeb and PatchDock.

## 4..1. pyDockWeb

pyDockWeb is a web server for the presumption of protein-protein interaction. It offers a web interface for rigid body docking and scoring. For running a project no registration is needed but it is advisable to enter a personal E-mail address to get a notification for available results.

The input of the protein structures can be in PDB file format. With the 3D coordinates of two interacting proteins, pyDockWEB evaluates the best docking locations, principally through electrostatics and desolvation energy. The output is a list

Due to an error when uploading the PDB-file for Ara h 3 the docking simulation could not be executed and therefore it was skipped. The table below shows the results of the antibody-antigen bindings iterating through all possible cases (exept Ara h 3) (PA12P3F10 was abbreviated to PA12, PA13P1H08 to PA13):

| Binding | Conf | Electrostatics | Desolvation | VdW | Total | Rank |
|---|---|---|---|---|---|---|
| PA12&AraH 1 | 8410 | -30717 | 1349 | 16722 | -27696 | 1 |
| PA12&AraH 2 | 2230 | -14250 | -12017 | 34149 | -22852 | 1 |
| PA13&AraH 1 | 5902 | -45270 | 11361 | 25704 | -31339 | 1 |
| PA13&AraH 2 | 4866 | -29793 | -0.127 | 65889 | -23332 | 1 |

Further results include a 3D simulation of the protein docking and a plot of the energy score. Shown below is the energy score of the antibody PA12P3F10 paired with the antigen Ara h 1:

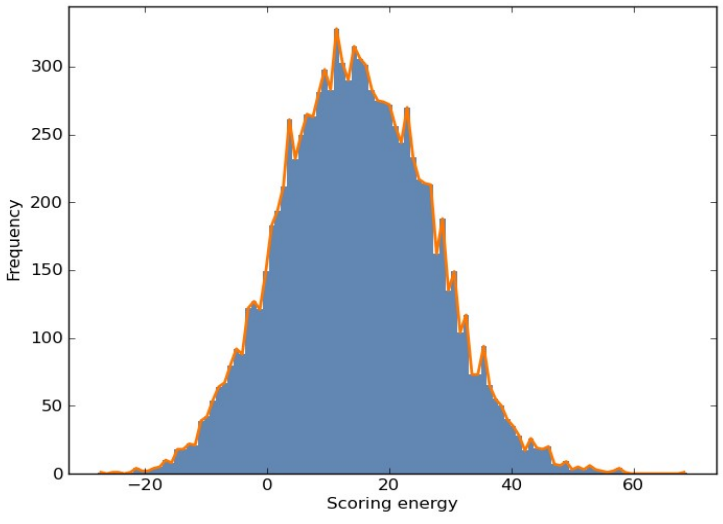


Fig 11. PA12P3F10 paired Ara h 1 scoring energy

Energy score of PA12P3F10 bound with Ara h 2:

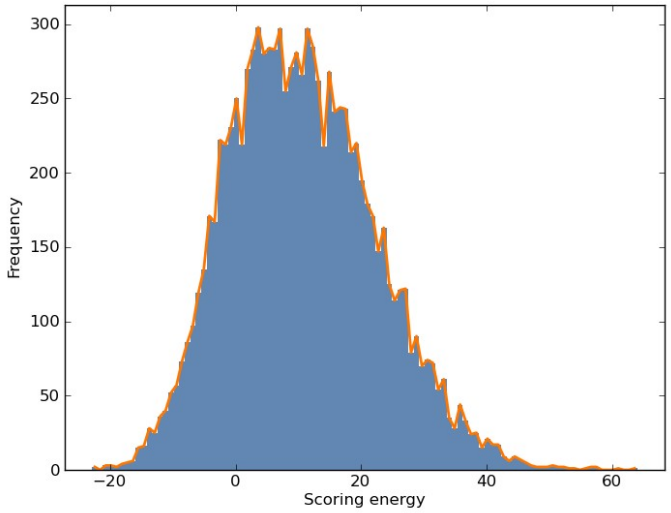



Fig. 12. PA12P3F10 bound with Ara h 2

Energy score of PA13P1H08 bound with Ara h 1:

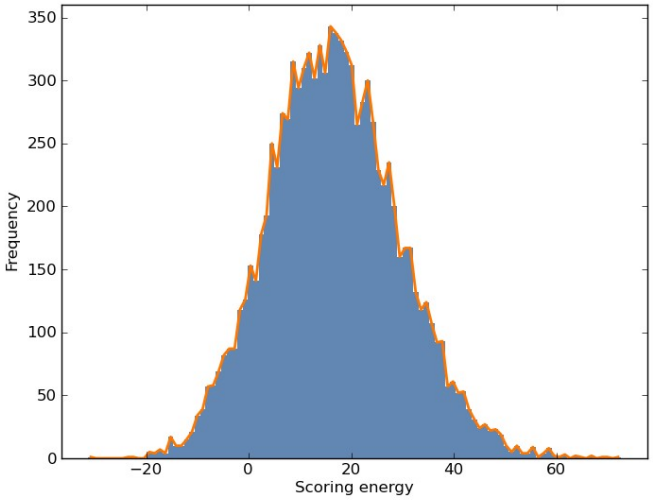



Fig. 13. PA13P1H08 bound with Ara h 1

Energy score of PA13P1H08 bound with Ara h 2:

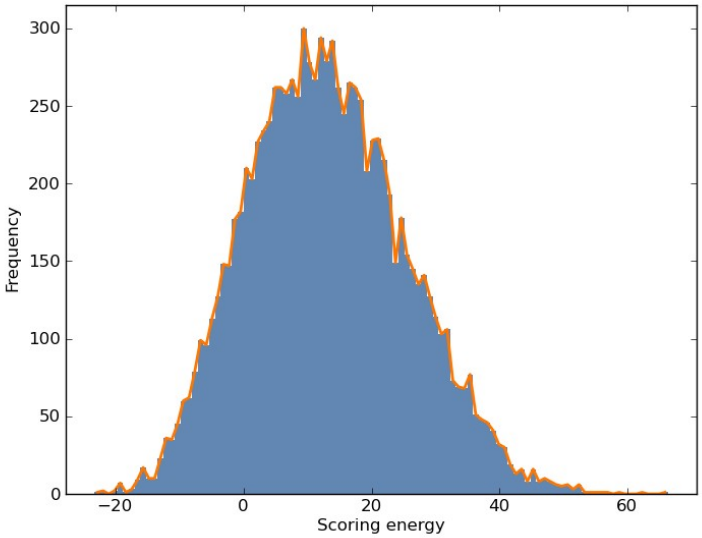



Fig. 14 PA13P1H08 bound with Ara h 2

## 4..2. PatchDock

PatchDock as well has a simple web interface which enables uploading local PDB files and being informed via e-mail when the result is calculated. The surfaces of two molecules are divided into surface form patches. These patches match the patterns which differentiate visually between puzzles. [ CITATION Sch05 \l 1031 ] The patches can be overlaid by using the shape matching algorithms once they are found. Three basic phases can be outlined:

- Molecular Shape Representation: The molecular surface of the molecule is measured in this step PatchDock. Next a geometric patch detection segmentation algorithm is used (concave, convex and flat surface pieces). The patches are filtered to only keep patches that have focal residues.
- Surface Patch Matching: In this step a fusion of geometric hashing and pose-clustering matches he patches detected in the molecular shape representation step. Concave patches pair with convex and flat patches go with any sort of patches.
- Filtering and Scoring: The complexes of the candidate are examined from the previous stage. With insufficient penetrations of the receptor atoms, it discards all the complexes to the ligand atoms. The other representatives are then evaluated according to a geometric complementarity rating.

The table below contains the results of all variations of the protein bindings (PA12P3F10 was abbreviated to PA12, PA13P1H08 to PA13):

| Binding | Solution No | Score | Area | ACE | Transformation |
|---|---|---|---|---|---|
| PA12&AraH 1 | 1 | 14820 | 2199.60 | 128.72 | 2.53 0.27 2.19 132.46 105.54 -5.39 |
| PA12&AraH 2 | 1 | 15776 | 2435.30 | 404.98 | -1.93 0.40 -2.43 48.67 83.50 11.88 |
| PA12&AraH 3 | 1 | 18258 | 2526.40 | 386.30 | -2.11 0.52 -2.45 54.71 114.64 -64.33 |
| PA13&AraH 1 | 1 | 14524 | 2301.90 | 439.66 | 0.47 0.24 2.69 55.43 2.98 84.31 |
| PA13&AraH 2 | 1 | 16508 | 2314.00 | 332.59 | -1.81 0.36 -2.48 51.45 81.70 15.84 |
| PA13&AraH 3 | 1 | 16724 | 2505.40 | 461.00 | 1.48 -0.28 -1.85 98.14 106.40 103.90 |

# 5. Discussion and Conclusion

In this work the investigated tools pyDockWeb and PatchDock were given 6 protein pairs (whereas Ara h 3 could not be uploaded to pyDockWeb due to structure size) to compute the antibody-antigen binding structure. The output parameters of both tools cannot be compared directly thus it is difficult to point out which is more accurate. For the binding of PA12P3F10 to AraH1 PyDockWeb exhibits an electrostatics value of -30717 and the desolvation lies at 1349. The Van der Waals column results in 16722. PatchDock delivers for the same proteins a score of 14820 and an ACE (Atomic contact energy) of 128,72, interestingly also one of the highest binding probabilities in contrast to the other protein bindings.

Compared to PA12P3F10 linked to Ara H 2 the pyDockWeb electrostatics are in the previous example about twice as high, which additionally can be interpreted as strong binding probability. The VdW returns a value of 34149. PatchDock delivered a score of 15776 and an ACE of 404,98. Thus, both tools predicted a lower antibody-antigen binding probability than in the first example.

Next the PA13P1H08 was bound to Ara H 1 and both tools agreed on a high binding probability. Pydock evaluated an electrostatics value of -45270 and a VdW of 25704. In relation, PatchDock provided a score of 14524 and ACE of 439,66.

Finally, the binding of PA12P3F10 with Ara H 2 dispensed with PyDockWeb a score number of -29793 and VdW of 65889, thus it figured out a low possibility for creating a bond. Likewise, PatchDock proceeded as well with a minor linking chance, gaining a score value of 16508 and ACE of 332,59.

As PatchDock was able to compute the antibodies with Ara h 3, the results as it follows are:

For PA12P3F10 bound to Ara H3 a score of 18258 and ACE value of 386,3 were recorded. These values conclude that a binding process of these two proteins is improbable.

Corresponding to the PA13P1H08 AraH3 docking the provided values score (16724) and ACE (461) also induced a slim chance for binding.

In the light of these results, it can be stated that the protein docking of PA13P1H08 with AraH1 returned from both tools the highest binding probabilities. As expected, the antibody docked onto the antigen with the most compatible structure of the variable domain. With PyDockWeb, the lowest equivalence of the fab-region turned out to be the binding of PA12P3F10 with AraH2, whereas PatchDock examined the linking of PA12P3F10 with AraH3. Unfortunately, this result cannot be compared to PyDockWeb because of the molecular size of AraH3.

Summing up, both tools delivered similar protein binding results and each of them has its advantages. However, as PyDockWeb has a limited uploading file size and both applications feature a similar computation time, PatchDock operated more effectively in this examination.

# Bibliography


[1] R. W. V. J. J. D. J. C. Y. A. Connie Rye, „OpenStax,“ Biology, Oct 21, 2016 Houston, Texas. [Online]. Available: https://openstax.org/books/biology/pages/3-4-proteins.

[2] „Scitable, Collaborative Learning Space for Science,“ natureEDUCATION, [Online]. Available: https://www.nature.com/scitable/ebooks/essentials-of-cell-biology-14749010/122996920/. [Zugriff am 14 12 2020].

[3] A. MacDonald, „www.technologynetworks.com,“ TECHNOLOGY Networks, 25 10 2017. [Online]. Available: https://www.technologynetworks.com/immunology/articles/antigen-vs-antibody-what-are-the-differences-293550#:~:text=Antigens%20are%20molecules%20capable%20of,response%20to%20exposure%20to%20antigens.. [Zugriff am 13 11 2020].

[4] K. Latham, „biologydictionary.net Editors, “Antigen vs. Antibody”,“ biology dictionary, 12 1 2021. [Online]. Available: https://biologydictionary.net/antigen-vs-antibody/. [Zugriff am 15 01 2020].

[5] T. E. o. E. Britannica, „Encyclopedia Britannica,“ Britannica, 27 5 2020. [Online]. Available: https://www.britannica.com/science/antibody. [Zugriff am 2 12 2020].

[6] "HADDOCK2.4 Antibody - Antigen tutorial using PDB-tools webserver," [Online]. Available: https://www.bonvinlab.org/education/HADDOCK24/HADDOCK24-antibody-antigen/. [Accessed 01 2021].

[7] K. E. K. a. A. A. Fodor, „arxiv.org,“ Department of Bioinformatics and Genomics, [Online]. Available: https://arxiv.org/ftp/arxiv/papers/1401/1401.1141.pdf. [Zugriff am 14 12 2020].

[8] J. P. G. L. M. R. D. D. V. H. A. M. J. J. B. Li C Xue, „Template-based protein–protein docking exploiting pairwise interfacial residue restraints,“ Briefings in Bioinformatics, 5 2017. [Online]. Available: https://academic.oup.com/bib/article/18/3/458/2562753. [Zugriff am 16 12 2020].

[9] „Protein Structure Prediction Center,“ [Online]. Available: https://predictioncenter.org/. [Zugriff am 5 1 2020].

[10] I. Y. N. R. W. H. Schneidman-Duhovny D, „PatchDock and SymmDock: servers for rigid and symmetric docking.,“ Nucleic Acids, 27 6 2005. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1160241/?tool=pubmed. [Zugriff am 20 1 2021].

[11] H. T. K. N. S. K. a. A.-U.-A. N. Kamakura, „Protein function prediction based on k-cores of interaction networks,“ 2010 International Conference on Bioinformatics and Biomedical Technology , pp. 211-215, doi:

10.1109/ICBBT.2010.5478977, Chengdu, 2010.

[13] P. C. F.-R. J. Jiménez-García B, „pyDockWEB: a web server for rigid-body protein-protein docking using electrostatics and desolvation scoring,“ Bioinformatics, 1 7 2013. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/23661696/#:~:text=pyDockWEB%20is%20a%20web%20server,of%20the%20pyDock%20scoring%20algorithm.&text=Given%20the%203D%20coordinates%20of,by%20electrostatics%20and%20desolvation%20energy.. [Zugriff am 7 1 2020].

[14] T. P. W. M. Janeway CA Jr, „Immunobiology: The Immune System in Health and Disease.,“ Garland Science, 2001. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK27160/. [Zugriff am 16 11 2020].

# List of Figures

YYYYYYYYYYYYYY

# List of Tables