Saturation Mutagenesis of the Z-domain

Degree project, in Biotechnology, first level

Karin Alexandersson Ros, Richard Brunsell, Matilda Fresk,
Alfred Kedhammar, Gabriella Rossi

May 12, 2019

Sammanfattning

Målet med projektet är att förbättra affiniteten mellan Immunoglobulin G (IgG) och den bindande Z-domänen hos protein A från stafylokocker genom att skapa ett bibliotek med muterade antikroppar. Z-domänen används inom bioteknik för många applikationer som skulle dra nytta av en ökad affinitet.

För att påverka affiniteten genomfördes en mättnadsmutagenes där alla ytexponerade aminosyror blir substituerade för alla andra aminosyror. Mättnadsbiblioteket bestående av 912 olika plasmider, där alla har en punktmutation, erhölls via en tredje part. Plasmiderna extraherades från *Escherichia coli* för att kunna ersätta nonsenssekvensen i plasmiden med en funktionell Z-domän. Innan ersättningen gjordes hade både plasmiden och den funktionella Z-domänen amplifierats via polymerase chain reaction (PCR). Plasmiderna sattes samman och transformerades in i *E.coli* för att odla upp produkt. Efter amplifiering i *E.coli* extraherades plasmiderna, renades och slutligen transformerades in i *Staphylococcus carnosus* för ytterligare amplifiering. Flödescytometri och cellsortering genomfördes för att analysera resultaten.

Affiniteten mellan IgG och Z-domänen är relativt högt i vildtypen, 20 nM [1], men genom mättnadsmutagenesen kommer affiniteten påverkas och förhoppningsvis förbättras. Given mängden av olika mutationer både en ökning och minskning av affinitet förväntas. Projektet har potential att identifiera punktmutationer och ge användbar information för framtida design av mättnasbibliotek.

Abstract

The project aims to enhance the affinity between the Immunoglobulin G (IgG) and the binding Z-domain from staphylococcal protein A by creating a library with mutated antibodies. The Z-domain is commonly used in biotechnology for many applications that would benefit from a high affinity.

To affect the affinity a saturation mutagenesis was performed where all surface-exposed residues are substituted for all of the different amino acids. The site saturation library consisting of 912 different plasmids, each with a single point mutation, was obtained via a third party. The plasmids where extracted from *Escherichia coli* in order to replace the dummy sequence in the plasmid with a functional Z-domain after having amplified both the plasmid and the insert with plymerase chain reaction (PCR). The plasmids where assembled and then transferred to *E.coli* to grow a lot of product, after having been amplified in *E.coli* the plasmids where extracted and purified to be able to transfer them into *Staphylococcus carnosus* for further amplification. Flow cytometry and cell sorting was performed to analyze the results.

The affinity between IgG and the Z-domain is relatively high in wild type, 20nM [1], but through the performed mutations the affinity will be affected and hopefully it has increased. Given the amount of different mutations both an increase and decrease in the affinity can be expected. The project has the potential to identify single-point mutations and give useful information for future designing of saturation libraries.

Contents

1 Background

1.1 Libraries and directed evolution

The ancient way of producing protein-based affinity reagents *in vivo*, involves injecting pathogens into the bloodstream of a laboratory animal, and then rely on the immune system and B-cells of the animal, to produce the monoclonal antibodies. The blood is later drawn from the animal to collect the desired protein-based affinity reagents matching the injected pathogen.

In vitro-based methods, such as directed evolution, phage display and library technology has been developed as an alternative to immunization of laboratory animals for *in vivo* protein-based lab experiments and screening for affinity reagents such as monoclonal antibodies. The method *directed evolution* won the Nobel Prize in Chemistry 2018 [2].

The above mentioned in *vitro methods* are based on altering a desired protein sequence by varying base pairs coding for the amino acids in a DNA-sequence. The desired protein without the modifications, is referred to as the wild type. The result is a library containing of millions or billions improved or worse versions of the wild type protein. This library is used to screen for the versions with the highest affinity for the specific pathogen. Thus, this methods use the same principles of antibody formation as the ancient way of producing protein-based affinity reagents but without the need for lab animals.

1.2 Affinity maturation

Affinity maturation is the engineering process of improving an already existing affinity protein to obtain an even higher affinity. The most forward approach to obtain this is to use an error-prone PCR to consciously introduce changes in the original DNA sequence to create a variation of different versions of the original protein [3]. The result of error prone PCR is often low quality libraries compared to site directed alternatives, with a majority of non functional proteins which is a result of induced new diversity in parts of the protein disconnected from the paratope and subsequently introduced deletions [4]. The folding of the protein is crucial and small changes in the amino acid structure can result in cardinal changes in the functionality of the protein.

1.3 Saturation mutagenesis, site directed mutagenesis and alanine scan

To analyze the function of an amino acid in relation to a whole protein site directed mutagenesis is often performed by doing an alanine scan. In an alanine scan a single surface-exposed residue in the paratope is substituted for alanine and the effect this has on the protein is analyzed [5].

A similar method that provides a more thorough analyze is site saturation mutagenesis which is when all surface-exposed residues are substituted for all of the different amino acids. This method provides a collection of mutagenesis products that all have a different amino acid in the target position, making the library so called saturated.

When producing a library through mutagenesis it is important to consider the desired randomness whilst being able to eliminate the parental DNA. When performing mutagenesis, most methods use an annealing primer, commonly performed by synthesizing it as a mixture of oligonucleotides that are identical to the parental DNA, only having a specific positions being non-identical. By creating a primer that is identical, allows the primer to anneal to the parental DNA whilst the non-identical positions creates the diversity of the product [6].

Site maturation mutagenesis is a simple and effective way for creating variants of proteins that would have taken a long time to occur naturally, or if even at all.

1.4 Immunoglobulin G

Immunoglobulin G is the most common and simplest antibody found in vertebrates consisting of two identical light chains and two identical heavy chains that are connected by disulfide bonds. An IgG antibody consists of two fragments, the antigen binding region (Fab) and the crystallizable region (Fc), where the Fab region contains the paratope which has the surface-exposed residues. The IgG antibody works to create a connection between the antigen and the body's effector system, the fab region binds to the antigen and the Fc region is connected to the effector system [7].

1.5 The Z-domain

There have been several bacterial immunoglobulin (Ig)-receptors found in later years, with Staphylococcal protein A (SpA) being well studied, and they have come to be important for the binding, identification and purification of immunoglobulins. Protein A was the first described bacterial immunoglobulin receptor and is found in the cell wall of Staphylococcus aureus, it has a molecular weight of 42 kDa and binds to the Fc region of IgG. By studying the structural changes that occurs when SpA binds to the Fc regions shows that this interaction involves the Z domain of SpA, studies have also shown that SpA can interact with the Fab region [8].

The SpA has become the foremost used reagent in immunological methods and a variety of applications have been developed. All methods that use SpA is based on the selective binding between the SpA and the Fc region of the IgG, these methods are divided into two categories. The first category involves methods that use purified SpA that have been labeled with a tracer group in order to detect and quantify antibodies and antigens. The second category uses immobilized SpA either as a part of the cell membrane of the bacteria or with purified SpA covalently bound to a solid support in order to isolate whole IgG or subclasses, to divide cells, selectively absorb IgG or to precipitate antibody-antigen complexes [9].

1.6 Polymerase chain reaction, PCR

Amplification with PCR is a common and effective method to generate a large number of copies of a desired DNA fragment. It is widely used because of its simplicity and ability to create product fast. To design primers for the PCR the program Benchling was used. The most prominent disadvantage with PCR amplification has to do with the primers needed. Since primers are necessary for this method, the sequence of the target region has to be known. Another disadvantage with this method the use of DNA polymerase which is error prone creating mutations in the generated product [10].

1.7 *In vitro* assembly

To construct a complete circular plasmid in vitro assembly was performed. This method uses exonucleases that chews back on the 5' ends of the fragments, leaving single stranded DNA regions. With the overlap created by the placement of the PCR primers, the ends of the plasmid and insert fragments can anneal, creating a complete plasmid. The regions where the exonuclease chewed back too far and left ssDNA outside the overlap is extended via DNA polymerase, and the nicks are sealed with cellular ligase *in vivo* or *in vitro* [1].

1.8 Cell sorting and flow cytometry

Flow cytometry is a technique used in analysis of cells and together with the cell sorting technique it is a powerful and efficient tool used in laboratory work with prokaryotic and eukaryotic cells *in vitro*. The method uses a combination of fluidics, optics and electronics to generate diagrams and mechanically sort viable cells one by one in different vials according to characteristics. The information obtained by this technique include measurement in size, cell count, cell cycle and characteristics. It is possible to obtain specific information on individual cells [11].

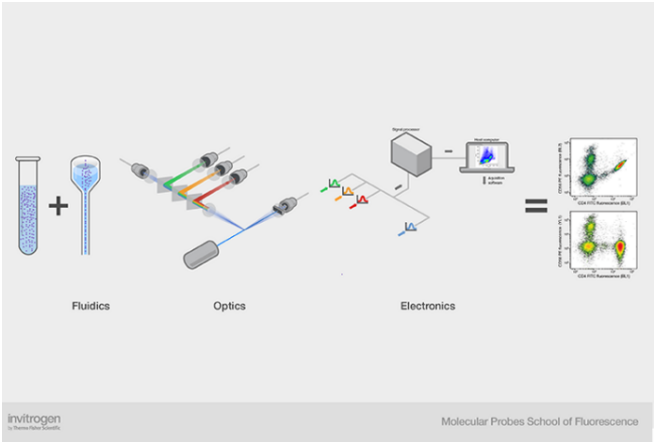


Figure 1: The different components that build up the basics of a flow cytometry and cell sorting. The combination of the different techniques creates a powerful and precise technique.

In this lab, protein A is expressed on the outer membrane of the *S. carnosus*. To examine the affinity for the mutated protein A, two antibodies with different fluorescent spectrum were used and incubated with the cells for the flow cytometry and cell sorting.

Since the mutations were specifically induced by saturation mutagenesis instead of a non specific method such as error prone PCR, it is possible to know which part of the protein is constant between the mutants. One of the antibodies bound to a region of protein A that were left non-mutated, this antibody measurement was used to count the amount of protein A on each cell. The second antibody bound to the mutated region of the protein, testing the affinity for that specific mutant. In combination of the two, information on the relative amount of protein A on each cell together with the amount of antibody binding to the mutated region, the affinity between the mutated protein A and Fc region on IgG was measured [1]. This information generated a plot diagram with one fluorescence on the x-axes and the other fluorescence on the y-axes, example in figure 1.

Cell sorting is then generated based on the information obtained by selecting gates in the plot diagram. Gates is marked areas in the plot diagram that is selected by the researcher for sorting. The flow cytometry is run once again and the computer will then sort out the cells that are gated in different tubes using electric fields. In this lab we are interested in sorting out two types of cells. One gate with high affinity between Fc region and protein A and another gate corresponding to lost affinity. The sorted cells can furthermore be sequenced to collect information on which mutations give higher affinity, and which mutations result in lower affinity.

When working with fluorophores, it is important not to subject them to light causing them to fade, which would make it hard to see any results, it is also important for

this method that the fluorophore has bound to the antibody. Whilst wanting to avoid having no signal it is not desired to have an excess fluorescent signal which is why the concentration is important, as the results show multiple concentrations where used to find what worked.

1.9 Overview of the procedure

The procedure will comprise affinity maturation of the IgG-binding Z-domain from Staphylococcal protein A. To create an enhanced affinity within the Z-domain, a saturation mutagenesis will be conducted.

An effective way to perform affinity maturation is to use targeted mutation, which means that the mutations are directed to exchange amino acids located on the surface of the paratope, thus be able to analyze the influence on their affinity. This can be done with alanine scanning where one replaces an amino acid residue on the surface of alanine and then analyze the effects on the affinity [1]. Alanine is the most simple of amino acids and will therefore silence out the effect of binding ability to that specific position in the protein. In this project inspiration is taken from the alanine scan but instead substitution of every other amino acid will be tested instead of only one single alanine. This will give a broader spectrum of variety off combinations of the protein and the possibility to find higher affinity, when alanine scan only provides information of silencing one position at the time. However the principles of alanine scan is used in the saturation mutagenesis of this project.

Saturation mutagenesis will be conducted in which mutants of selected amino acids are mutated by replacing them with all different amino acids to see the influence on the affinity. By doing this, results from all possible substitutions, not just alanine, will be collected. This project will be limited to individual point mutations in order to examine the impact of specific mutations on the protein. Once the effects of the mutations have been observed in a flow cytometer, the data will be analyzed. The individual point mutations will give 912 different variants of the Z domain and when the effect of the individual mutations has been studied they can be combined to create an even more targeted library [1].

2 Aims of the project

The aim of this project at large is to generate, interpret and present data regarding how all possible point substitutions of 19 amino acids in the paratopes of the Z-domain affect it's affinity for the human antibody immunoglobulin G domain fragment crystallizable to facilitate the development of a targeted library for further affinity maturation.

In further detail, this aim amounts to a primary goal of producing a library consisting of 912 mutated variants of the Z-domain (19 substitutions on 48 sites respectively), each coding for a unique Z-domain. The next interim is to assemble an aliquota of this library to a suitable vector allowing for selective growth and surface expression on a bacterial cell. The vector is then transformed into competent cells to form the in-vivo library, which can be amplified or stored on demand. The affinity of the library can then be analyzed and compared to the wild type using fluorescence activated cell sorting (FACS), the results of which can be traced back to it's originating substitution using DNA-sequencing.

3 Project plan and methodology

3.1 Library construction

The complete library consisted of 912 different plasmids, each with a single point mutation (see appendix A).

The mutation was at one of the 48 predetermined positions of the Z-domain of Protein A, chosen so that the mutation affects the surface of the protein and thus potentially the relative binding to IgG Fc. The point mutations were substitutions of the original amino acid to any of the other 19 amino acids not present in the wild-type variant of Protein A's Z-domain.

Oligonucleotides of the mutated Z-domain were produced and sequence verified via a third party, Twist Bioscience (specification: site saturation library). The specified library order arrived in a 96-well plate. There were 48 wells, one for each altered position and each containing 19 variants of amino acid substitutions in that position.

The wells each contained approximately 50 ng of dry freeze material, these were diluted in 100 μL MQ to an approximate concentration of 0.5 ng/μL.

3.2 Library vector

The plasmid pSCZ1 [12] was used as a vector for the library. This plasmid contains the necessary expression sequences such as ORI, antibiotic resistance genes, and the surface expression and detection protein coding sequences. However, pSCZ1 does not contain a functional Z-domain but a dummy sequence in place. The dummy sequence is filled with nonsense and stop codons in different reading frames to ensure it does not interfere with any desired plasmid, should it be expressed. See appendix C.

These plasmids were extracted and purified from RR1 *E.coli* using a QIAquick Plasmid Purification Kit (QIAGEN). A gel electrophoresis was performed to ensure that the purification was successful.

3.3 Amplification of plasmid and insert via PCR

The plasmid pSCZ1 and the insert library were amplified by Polymerase Chain Reaction (PCR), according to PCR Protocol for Phusion ®High-Fidelity DNA Polymerase (M0530) at 50 μL (New England Biolabs). The 49 reactions (48 variants + negative control) performed on the insert library used HF buffer and the 8 reactions (7 parallel reactions + negative control) performed on pSCZ1 used GC buffer.

The PCR protocol for both procedures was derived using the NEB Tm calculator. A gel electrophoresis was performed to verify product.

Primers for both amplicons were designed using *Benchling*^{TM} to fit the recommendations of abovementioned protocol and synthesized by Integrated DNA Technologies, BVBA. The insert primers were designed so that the Z-domain and an overhang of 200bp were included. The vector primers excluded the innate nonfunctional dummy sequence, and had a similar overlap of 200bp. See appendix C.

The PCR-product was verified using gel electrophoresis. Original plasmid template in the pSCZ1 PCR product was destroyed using DPNI. All insert reactions and vector reactions were pooled and purified using QIAquick PCR Purification Kit (QIAGEN).

The purification was verified using gel electrophoresis.

3.4 Assembly

The PCR amplification of the vector created a linear construct. This fragment and the insert had an homologous region of 200bp at each end. *In vitro* assembly was used to assemble the linear constructs into a complete, circular, plasmid.

The assembly was carried out using two different kits, NEBuilder HiFi DNA Assembly Kit, and In-Fusion HD Cloning Kit. In the NEBuilder Kit, 100 ng of vector and 5.76 ng of insert were added to NEBuilder HiFi DNA Assembly Master Mix to give a ratio of vector to insert of 1:3.

The amount of NEBuilder HiFi DNA Assembly Master Mix added to the cells for the transformation was varied at 2, 3 and 5 μL respectively to evaluate which was the most efficient method. For the In-Fusion HD Cloning Kit the volume was kept constant at 2.5 .

In the In-Fusion Kit assembly, 100 ng of vector and 14.65 ng of insert were added to 5X In-Fusion HD Enzyme Premix to a ratio of vector to insert of 1:5. These mixtures were incubated at 50^{o}C for 15 minutes after which they were transformed to *E.coli*.

The transformation to *E.coli* was performed according to respective kit protocols, however the following deviations were made. The In-Fusion protocol calls for a washing step which was omitted. All the reaction mixture from both kits were used and spread on plates.

As an initial trial of the different kits, each kit was used in triplicate, with one negative control.

3.5 Amplification in *E.coli*

The assembled plasmids were transformed to*E.coli* via heat shock at 48^{o}C and grown on ampicillin plates overnight at 37^{o}C. The assembled plasmid contains the gene for beta-lactamase to ensure selective growth.

The number of colonies were counted and used to calculate library coverage using summarized probability and a negative exponential distribution formula.

The plasmids were harvested from the *E.coli* cultivation with a QIAquick Maxi Plasmid Purification Kit (QIAGEN).

3.6 Transformation and amplification in *S.carnosus*

The amplified and purified vectors from the *E.coli* cultivation were transformed into electrocompetent *S.carnosus* using electroporation.

Plasmid and 240 μL electrocompetent cells were thawed on ice for 5 min. The cells were then incubated at room temperature (RT) for 30 min, mixed and spun down. The cells were then heat-shocked in a water-filled heat block at 56^{o}C for precisely 1 min 30 seconds, after which 1 mL 0.5 M sucrose + 10% glycerol were immediately added to the tube which was then mixed by inversion.

The tube was then centrifuged at 4500x g (7000 RPM) for 10 min at RT. The supernatant was discarded. After this, 140 μL 0.5 M sucrose + 10% glycerol was added and the sample was mixed by pipette. Subsequently, 4 μL plasmid was added using a 10 μL filtered pipette tip, the tube was then mixed by pipette and spun down. The tube was incubated at RT for 10 minutes.

A BioRad MicroPulser was used for the electroporation with voltage 2.3 kV and time constant 1.1 ms.

The contents of the tube were evenly divided between 4 parallel electroporations, as follows. From the cell suspension, 55 μL were added to a cuvette using a pipette

and a 100 μL filter tip. The cuvette was then tapped into the bench to get rid of air bubbles. To the cuvette, 1 mL of B2 medium was added and the electroporation was performed. Immediately after, an additional 400 μL B2 medium was added, the sample was then mixed by pipette and transferred to an eppendorf tube.

The cells were then incubated at 37^{o}C under rotation for 2 hours after which 200 μL from each suspension was plated on RT-incubated growth plates with 10 ug/mL chloramphenicol at 37^{o}C for 24 hours. The cells were then incubated at RT for an additional 3 days.

The colonies on the plates were counted and library coverage was calculated again.

For amplification, the cells on the plates were transferred to a solution of Tryptic Soy Broth with yeast extract (TSB+Y) and chloramphenicol and grown at 37^{o}C for 24 hours.

3.7 Flow cytometry and cell sorting

A flow cytometry was performed on the library and the wild-type protein to determine the optimal concentration of IgG. From overnight cultures 15 μL cells were washed with 800 μL phosphate buffered saline with 0.1% Pluronic (PBSP). The cells were centrifuged at 6000 RPM in 4^{o}C for 6 min and the supernatant was discarded. The resulting cell pellets were resuspended in 250 μL PBSP containing 0.4, 1, 4, 20 and 100 nM biotinylated IgG. The cells were incubated at RT under gentle mixing for 45 minutes. The cells were washed with 180 ice-cold PBSP, and resuspended in 230 ice-cold PBSP containing the fluorophores HSA-Alexa-647 conjugate and Streptavidin Phycoerythrin-conjugate (SAPE) at a 1:500 dilution. The cells were then incubated on ice and in the dark for 45 min, after which they were washed with 180 ice-cold PBSP and resuspended in 300 ice-cold PBSP in flow cytometer tubes. The ratio of fluorescence at 580 and 665 nm was measured.

For cell sorting, the procedure above was repeated using a FACS machine with IgG concentrations 0.4 and 10 nM. The cells with higher affinity for IgG than the majority of binding cells were sorted at IgG concentrations of 0.4 and 10 nM. The cells with average binding affinity to IgG were sorted at IgG concentrations of 10 nM. The non-binding cells were sorted at IgG concentration 10 nm. The affinity gates were combined with a gate comparing front- and backscatter to only sort cells with a profile similar to *Staphylococcus*. The cells from the gates were grown on separate plates with chloramphenicol in 37^{o}C overnight.

3.8 Sequence analysis

From each plate 96 colonies were individually picked and amplified with PCR. To verify that the PCR was successful 10 colonies from each PCR plate were picked and analysed with gel electrophoresis. The amplified colonies were sent to Microsynth Seqlab Sequencing for bi-directional sequencing with primers SAPA-23 and SAPA-24 (see appendix C).

The returned sequencing results were obtained in the form of .fasta and .ab1 files for each primer. The .fasta files were converted to .csv using console commands. Next, a Matlab script was constructed (see appendix E). The script compiled all SAPA-23 sequencing results to a data vector, then it used sequences 20 bp up- and downstream of the substituted section of the Z-domain to align them in the correct reading frame. Next it used open-source software to translate the aligned nt sequences to their corresponding aa sequences. It then compared these sequences to the aa sequence of the wild type Z-domain and subsequently found the positions where AA substitution had taken place and in what qualitative form. During this process, the script also continuously classified and documented different discrepancies in the sequencing results that would prevent them from generating the desired results.

4 Results

4.1 Library construction

The library was delivered as specified in section 3.1.

4.2 Library vector

The successful extraction and purification of pSCZ1 from RR1 *E.coli* was verified with gel electrophoresis, see figure 2. The ladder to the left is used to verify the length of the purified fragment. Since the latter contains a known composition of fragments with known lengths, it is possible to verify the successful extraction. Gel electrophoresis is a time effective, fast, simple and relative cost efficient way to achieve results that are sufficient enough to continue with the vector to the next step.

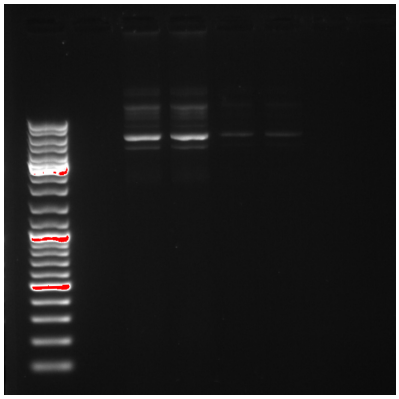
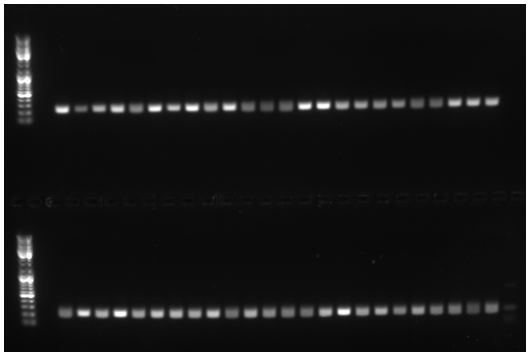


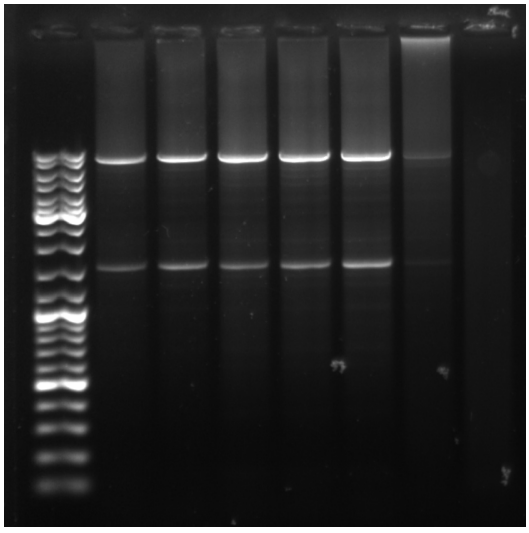
Figure 2: Gel verifying the extraction of pSCZ1 from RR1 *E.coli*.

4.3 Amplification of plasmid and insert

After the plasmid, pSCZ1, had been extracted a PCR was performed to amplify both the plasmid and the delivered inserts to create many copies. The PCR-product was verified using gel electrophoresis, see figure 3.

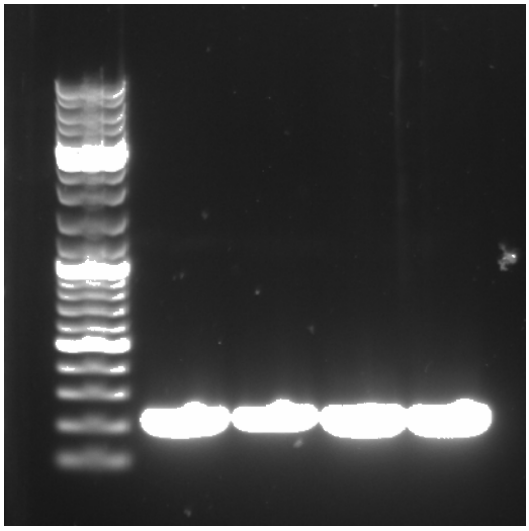


(a) Library amplicon.

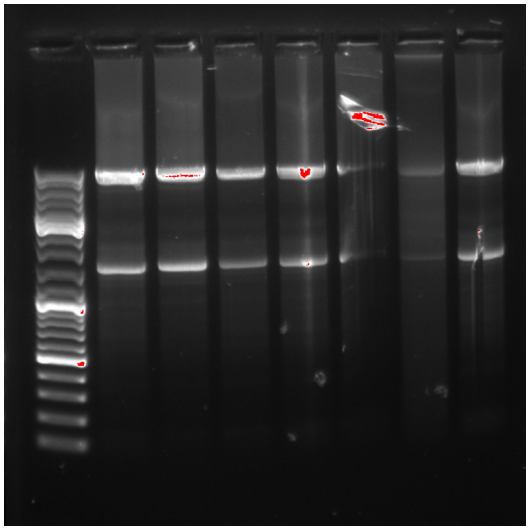


(b) Plasmid amplicon.

Figure 3: Gels verifying successful PCR-amplification.



(a) Library purification.



(b) Plasmid purification.

Figure 4: Gels verifying successful PCR-amplification.

The subsequent purification was verified using gel electrophoresis, see figure 4.

4.4 Assembly

The concentrations of the plasmid and the inserts were determined with nanodrop (see appendix). As the gel electrophoresis of the plasmid displayed a potential incomplete product in addition to the expected product, two thirds of the measured concentration was assumed to be the complete plasmid product, while the incomplete sample was ignored as it would not have the overlapping regions to proceed through assembly.

Table 1: Concentrations obtained via nanodrop. Sample 6 is omitted from the mean, and corresponds to well 6 in figure 4b

Concentration ng/						
89.2	85.3	73.9	79.3	78.9	44.5	73.9
Mean						**419**

A preliminary transformation was performed, and after an overnight cultivation the NEBuilder kit was determined to be of interest for continued projects as the growth was tenfold. The remaining of the purified fragments were assembled the following day. The NEBuilder reaction mixture was performed in quintuplicate of the three different concentrations yielding a total of 15 growth plates, while the In-Fusion mixture was plated on 3 growth plates.

Table 2: Table of colony count on respective growth plate from different assemblies

Kit	Colony count			Avg.	Neg. control
In-Fusion	199	168	196	187.6	1
NEBuilder	1088	1136	971	1065	22

Successful transformation was indicated by growth, as only a complete plasmid can express the ampicillin resistance gene, and allow the cells to survive on ampicillin growth plates.

4.5 Amplification via *E.coli*

The transformation to *E.coli* was successful for both NEBulider and In-Fusion and growth of colonies could be observed. The colonies were counted and the number of colonies can be seen in tables 3 and 4 below.

The concentration of reaction mixture per plate was varied when using the NEBuilder. That resulted in a difference in growth which can be seen in the average number of colonies per plate in the table below. The total sum of colonies were approximated to 21 678 colonies when using NEBulider and 3468 colonies when using In-Fusion. The grand total was approximated to be 25 146 colonies.

Table 3: Colony count using the NEBuilder kit

Vol.	Colony count						Avg.
5 ul	832	304	348	105	135	2808	755
3 ul	1616	1820	542	1212	3512	3217	1986.5
2 ul	2824	1220	1188				1744
Sum							**21678**

Table 4: Colony count using the In-Fusion kit

Vol.	Colony count			Avg.
2.5 ul	1456	652	1360	1156
Sum				**3468**

A calculation was made to quantify the probability of all variants of the library being present in the transformed cells. The result of the calculation strongly indicated that this was the case. See appendix D.

4.6 Transformation and amplification in *S.carnosus*

The transformed *S.carnosus* grew overnight on chloramphenicol plates generating new colonies with the mutated plasmid in the bacteria. The obtained colonies are assumed to contain the mutated plasmid based on antibiotic resistance gene present in the plasmid.

The colonies were counted on the plate grown with diluted cell suspension. The total number of colonies could then be calculated.

From the original bacterial solution:

$$4.5ml \xrightarrow{10\mu L} 1mL \xrightarrow{100\mu L} 25\ colonies \tag{1}$$

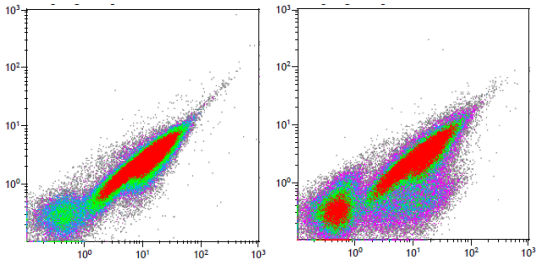
Implies a dilution factor of

$$25 * \frac{1mL}{10\mu L} * \frac{4.5mL}{10\mu L} = 112500 \tag{2}$$

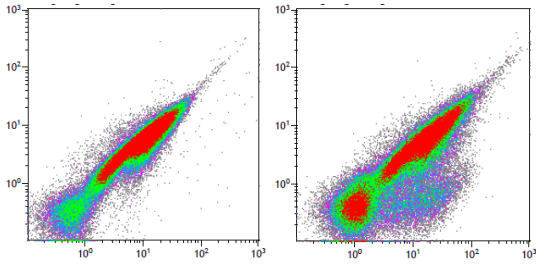
The total number of CFU in the original solution is thus calculated to 112 500.

4.7 Flow cytometry and cell sorting

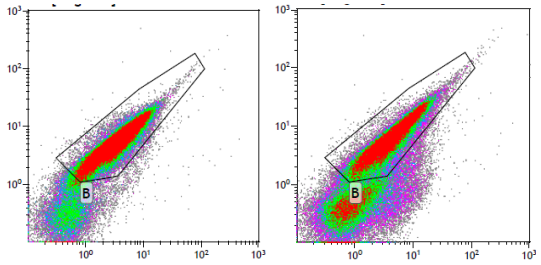
Flow cytometry was performed with six different concentrations in total, having done one procedure and then lowering the concentration for more optimal results. The first time the concentrations 100, 20 and 4 nM where used and as can be seen in figures 5e to 5j a certain decrease in affinity can be seen in the library results. The concentrations where lowered to 4, 0.4 and 0.1 nM for the second flow cytometry to further accentuate this. In figures 5a to 5d it is easier to see the decreased affinity in comparison to the original non-mutated protein A.



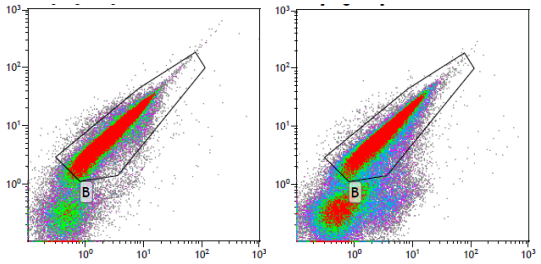
(a) Lib. at 0.4 nm IgG. (b) wt at 0.4 nm IgG.



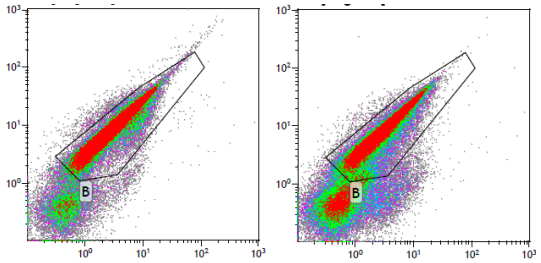
(c) Lib. at 1 nm IgG. (d) wt at 1 nm IgG.



(e) Lib. at 4 nm IgG. (f) wt at 4 nm IgG.



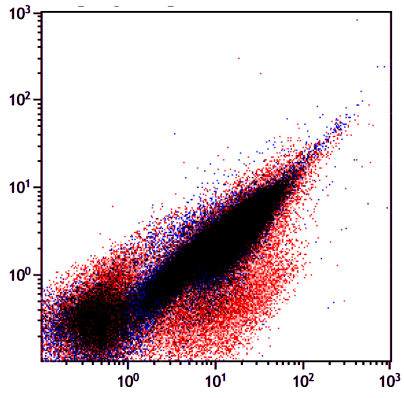
(g) Lib. at 20 nm IgG. (h) wt at 20 nm IgG.



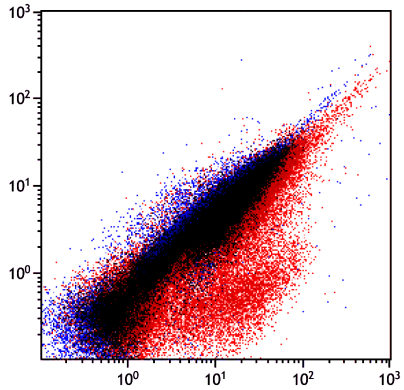
(i) Lib. at 100 nm IgG. (j) wt at 100 nm IgG.

Figure 5: Density dot plot of the *S.carnosus* cells with the library proteins and with wild type protein analyzed with flow cytometry. Five different concentrations of IgG are represented (0.4, 1, 4, 20 and 100 nm). The y-axis is the IgG binding signal and the x-axis is the surface expression level.

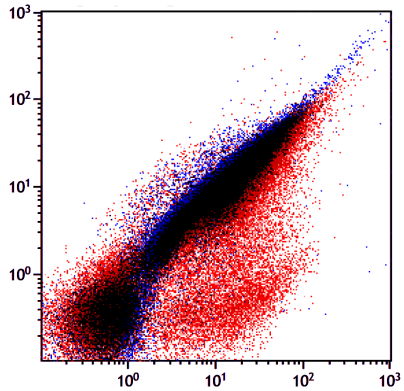
In figure 6 the results for the mutated plasmid and the wild type have been laid on top of each other to show the differences in affinity, the mutated plasmids are shown in red and the wild type in blue. This pictures clearly shows that the wild type have an overall higher affinity.



(a) 0.4 nm IgG.



(b) 1 nM IgG.

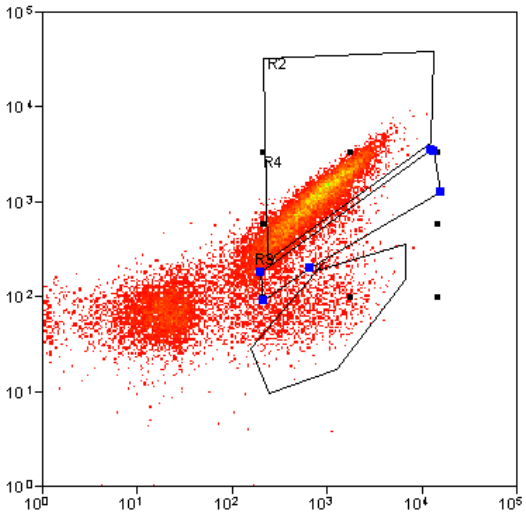


(c) 4 nM IgG.

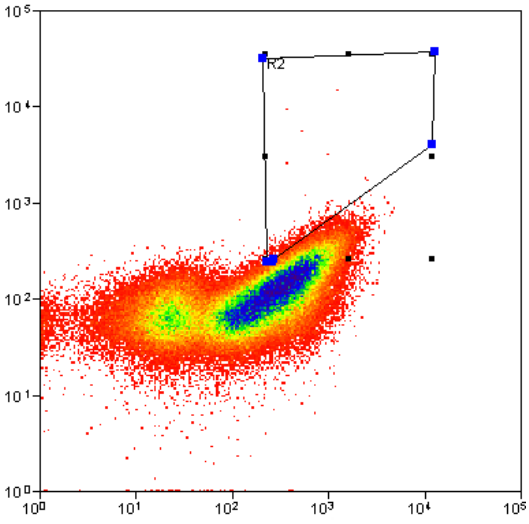
Figure 6: The density dot plots from the flow cytometry analysis of the *S.carnosus* cells with the library and wild type overlayed. The cells with the library proteins are represented in red and the cells with the wild type protein are represented in blue. The y-axis is the IgG binding signal and the x-axis is the surface expression level.

After having performed flow cytometry and found concentrations that worked, a FACS was performed to sort the cells based on affinity. The sorting was based on

three gates dividing the cells into high, medium and low affinity. Although no sorting was performed in the wild type for comparison, since the other results showed that the cells with the mutated plasmids had lower affinity, the same is expected here. As seen in the pictures most of the cells are in the gate for high affinity.



(a) 10 nM IgG.



(b) 0.4 nM IgG.

Figure 7: Density dot plot of the *S.carnosus* cells with the library proteins analyzed and sorted into gates with flow cytometry. The y-axis is the IgG binding signal and the x-axis is the surface expression level.

4.8 Sequence analysis

The final output of the script was the names of the sequencing wells, the position of substitution, the corresponding aa of the wild type and finally the corresponding aa of the variant, all arranged in columns and sorted by ascending position.

This data can be found in appendix F.

Comments denoting discrepancies (if any), the full sequencing results, the aligned sequences and their corresponding aa sequences can also be found in the full results, which are not disclosed in this report for spacing reasons.

5 Discussion

The interpretation of the result is based on the gates from the plot diagrams in the FACS analysis. Sorting was performed and then 96 samples from each of the four gates were collected. From these sequenced samples, we have detected what specific mutations were present in each of the four groups. Data regarding all four gates is found in appendix F.

When comparing the flow cytometry results from the wild type protein to the library, no definite conclusion can be drawn. The only definite conclusion is, when altering the protein many variants have got a lower affinity than the wild type protein. There are some library proteins that display a similar affinity as the wild type but not any that display a clearly higher affinity. It can be debated whether strains with higher affinities than wild type are present, and subsequent flow cytometry with pure strains of selected mutants could be used to determine relative affinity of a specific mutant. This could be investigated in a directed evolution experiment where subsequent FACS and growth of the library sift through high binders.

Conclusions that can be drawn from the data available from one cell sorting are based on the DNA sequences obtained from sequencing. Information about relative mutation frequency of different amino acid positions can be determined. The observed mutation frequency of a position in a affinity category relates to the effect said mutation has on the binding affinity. This is presented in figure 8.

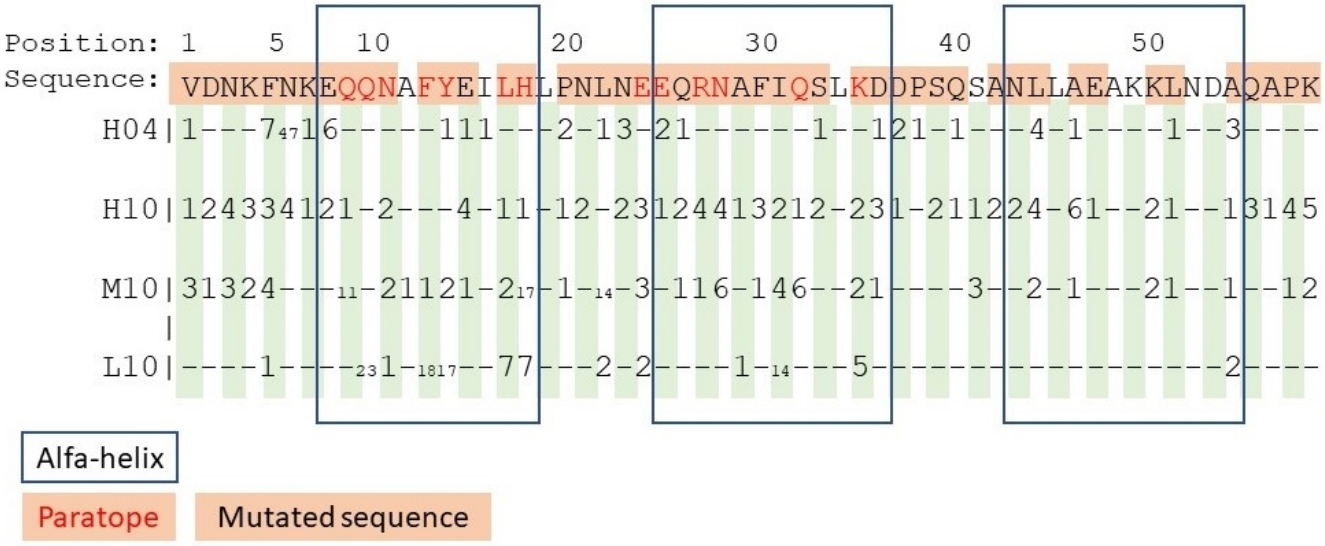


Figure 8: The numbers represent the numbers of mutations per position in the different gates (H04, H10, M10 and L10). The paratope is represented in red and the mutated sequence is represented with orange. The alpha-helix is marked with black squares.

Position 10, 13, 14 och 31 were mutated in more than 14% of the 96 cells taken from the gate with low affinity, at 10 nM IgG (L10). These positions are all located in the A and B alpha helix of the Z-domain. Position 10, 13 and 14 are part of the 13 positions initially suggested for mutation, and can be assumed to be amino acids displayed on the part of the surface that interacts with IgG and are therefore significant to affinity.

Position 6 was mutated in almost 50% of the cells with high affinity at 0.4nM IgG (H04), implying either great importance of the amino acid in that position, or no importance at all. As the high affinity binders had a similar affinity to Zwt, it can be assumed that the affinity is no better than Zwt. The mutation at position 6 presumably gives the same binding to IgG as the wild type and carries little importance. This is supported by the sequences (see appendix F) which have mutations to 14 of the 20 different amino acids possible, and the fact that it is located outside any alpha helix.

Position 9, 18 and 22 were highly mutated in the region of low affinity binders, but not non-binding to IgG at 10 nM IgG (M10). These positions were mutated to many different amino acids, which presumably affected and lowered the binding affinity. Two of these positions are located in alpha helix A, and included in the 13 positions of interest. It can be assumed that these positions are important for binding to IgG, but not essential as many different amino acids can be substituted and still provide affinity.

The exploratory nature of this project offered an opportunity to explore all results and what information they could provide and not just be focused on one specific outcome. Since the affinity between IgG and protein A already are used in the industry, there is interest in developing an improved affinity. The result from this project can therefore be used as a stepping stone to start exploring that possibility, even though this project was unable to do so. The project has the potential to identify single-point mutations and give useful information for future designing of saturation libraries.

The investigative nature of the project could also have been a threat to the project as there was no certainty as to whether or what result would be obtained. The time constraints on the project was also a factor in the results. If more time had been available more thorough investigations could have been made with a larger sample size, giving the possibility to alter the experiments to see if the results could be improved. Another possibility is investigating the folding and affinity of the protein *in silico* and compare to the *in vivo* results.

Nowadays, the Z-domain is used in medical and biotechnological applications. Some of these applications and processes would be improved and made more efficient by a better bond, thereby benefiting from the results obtained by the project. An example of an application that can benefit from a higher affinity is the development of drugs with the use on antibodies, the application is based on a process where the Z domain can be used to immobilize or extract antibodies without affecting their active site. A better binding of the Z domain could here result in a higher yield of antibodies.

The pharmaceutical industry has a large turnover and a high economical yield on development. In the end it is possible to reduce the cost of pharmaceuticals. This can affect the world on a global scale, as drugs can be developed more easily, and become cheaper, making them

more accessible and more able for society to benefit from. Higher yields in production stages also contribute to less wastage and lower emissions to the environment, or required resources in waste treatment. By doing experiments like this and continuing to explore possibilities for increased affinity, ultimately information obtained from such experiments could affect the problems described here.

The optimal consequence of this project would be to influence humans and society in a positive way, giving people a healthier and higher quality of life. The experiment could also provide information beneficial for the environment, if more protein and antibody-based drugs were to be used. Today, most drugs available on the market are chemical-based [13], they can be difficult to break down when the molecules leave the body [14]. In many cases, these chemicals have a negative impact on the environment, water quality and ecosystems. Protein-based drugs are developed based on protein found naturally in the body, IgG, this makes them easier to break down.

References

[1] Löfblom J. Personal communication with project advisor; 2019.

[2] The Nobel Prize in Chemistry 2018;. Available from: https://www.nobelprize.org/prizes/chemistry/2018/popular-information/.

[3] Mishra AK, Mariuzza RA. Insights into the Structural Basis of Antibody Affinity Maturation from Next-Generation Sequencing. Frontiers in Immunology. 2018 Feb;9:117. Available from: http://journal.frontiersin.org/article/10.3389/fimmu.2018.00117/full.

[4] McCullum EO, Williams BAR, Zhang J, Chaput JC. Random mutagenesis by error-prone PCR. Methods in Molecular Biology (Clifton, NJ). 2010;634:103–109.

[5] Morrison KL, Weiss GA. Combinatorial alanine-scanning. Current Opinion in Chemical Biology. 2001 Jun;5(3):302–307. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1367593100002064.

[6] Siloto RMP, Weselake RJ. Site saturation mutagenesis: Methods and applications in protein engineering. Biocatalysis and Agricultural Biotechnology. 2012 Jul;1(3):181–189. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1878818112000734.

[7] Sanders PW. Chapter 24 - Pathogenesis of Paraproteinemic Renal Disease. In: Mount DB, Pollak MR, editors. Molecular and Genetic Basis of Renal Disease. Philadelphia: W.B. Saunders; 2008. p. 435–440. Available from: http://www.sciencedirect.com/science/article/pii/B9781416002529500288.

[8] Justiz-Vaillant A, McFarlane-Anderson N, Smikle M. Bacterial Immunoglobulin (Ig)-Receptors: Past and Present Perspectives. American Journal of Microbiological Research;p. 7.

[9] Langone JJ. Applications of immobilized protein A in immunochemical techniques. Journal of Immunological Methods. 1982 Dec;55(3):277–296. Available from: https://linkinghub.elsevier.com/retrieve/pii/0022175982900886.

[10] Li CM, Dong H, Zhou Q, Goh KH. CHAPTER 11 - Biochips – fundamentals and applications. In: Zhang X, Ju H, Wang J, editors. Electrochemical Sensors, Biosensors and their Biomedical Applications. San Diego: Academic Press; 2008. p. 307–383. Available from: http://www.sciencedirect.com/science/article/pii/B9780123737380500131.

[11] Introduction to flow cytometry;. Available from: https://www.abcam.com/protocols/introduction-to-flow-cytometry.

[12] Fleetwood F, Klint S, Hanze M, Gunneriusson E, Frejd FY, Ståhl S, et al. Simultaneous targeting of two ligand-binding sites on VEGFR2 using biparatopic Affibody molecules results in dramatically improved affinity. Scientific Reports. 2015 May;4(1):7518. Available from: http://www.nature.com/articles/srep07518.

[13] IFPMA Facts and Figures Report – IFPMA;. Available from: https://www.ifpma.org/resource-centre/ifpma-facts-and-figures-report/.

[14] Pharmaceuticals in the environment: a growing problem;p. 8.

A Library

Wild type aa sequence:
KFTSQASDKKPTPRSAQDPLEVDNKFNKEQQNAFYEILHLPNLNEEQRNAFIQSLKDDP
SQSANLLAEAKKLNDAQAPKVDLQACKLLDALAKAKADALKEFNKYGVSDYYKNLINN

Length:
117

Substituted positions:
22-32, 34-36, 38-39, 41-54, 56-61, 63-65, 67-68, 71-72

Total number of substituted positions:
48

B Plasmid pSCCZ1 composition

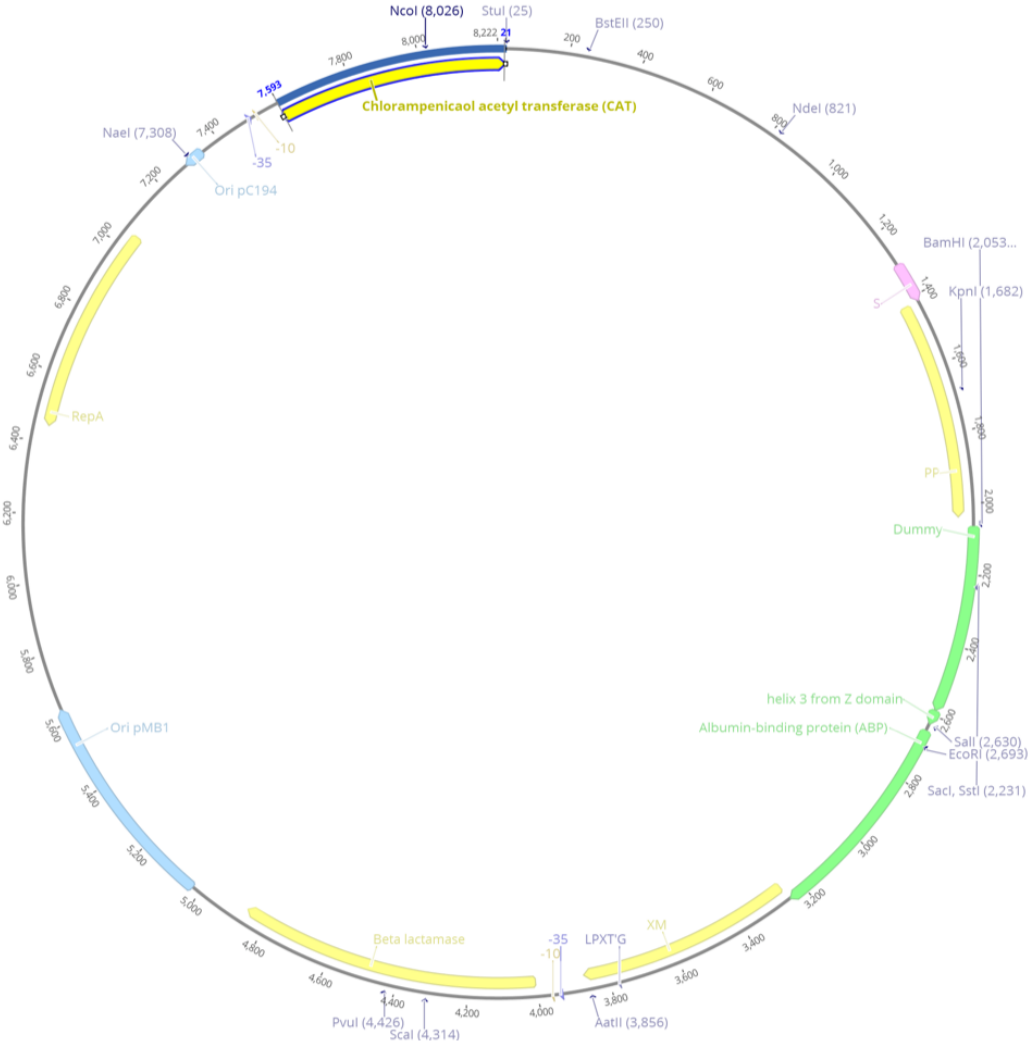


Figure 9: Circular map showing the components of the plasmid pSCZ1.

C Primer sequences

Table 5: Primers used for amplification of variants and plasmid for assembly.

Denotation	**Sequence**
Plasmid forward	caggcatgcaagcttctagatgctc
Plasmid reverse	tcctgcgcagatcttggtgttg
Insert forward	agatctgcgcaggatcccct
Insert reverse	aagcttgcatgcctgcaggtc
SAPA23	?
SAPA24	?

D Calculation of library coverage

Let n be the number of variants 912 and N the number of transformed colonies 25146.

D.1 Calculation using standard fractional probabilities

The calculation uses the following assumptions:

- All variants are present at equal concentrations in the reaction mixture
- All variants are equally likely of being transformed
- All transformed variants are equally likely of growing
- The odds of encountering a particular variant is independent of the number of times that same variant has already been encountered

Hence, the probability of encountering a particular variant in one colony is

$$\frac{1}{n} \tag{3}$$

the probability of not encountering a particular variant in one colony is

$$1 - \frac{1}{n} \tag{4}$$

the probability of not encountering a particular variant in at least one of all colonies is

$$(1 - \frac{1}{n})^N \tag{5}$$

the probability of encountering a particular variant in at least one of all colonies is

$$1 - (1 - \frac{1}{n})^N \tag{6}$$

the probability p of encountering all particular variants in at least one of all colonies is

$$p = (1 - (1 - \frac{1}{n})^N)^n \tag{7}$$

corresponding to a probability of 99.9999991%.

D.2 Calculation using negative exponential distribution formula

$$p = 1 - e^{-\frac{N}{n}} \approx 1 \tag{8}$$

E Matlab script for sequence analysis

```
%group = 'L';

groupvector = cell(4,1);
groupvector{1} = 'H04';
groupvector{2} = 'H10';
groupvector{3} = 'M';
groupvector{4} = 'L';

for groupcounter = 1:4
group = groupvector{groupcounter};

%Generate file names
Alphabet = 'ABCDEFGHIJKLMNOPQRSTUVWXYZ';

filenames_letters = num2str(zeros(96,1));
filenames_numbers = zeros(96,1);
letter = 1;
for i = 1:12:96
filenames_letters(i:i+11) = Alphabet(letter);
letter = letter+1;
for j = 0:11
filenames_numbers(i+j) = j+1;
end
end
filenames_numbers = num2str(filenames_numbers);

for i = 1:96
if str2num(filenames_numbers(i,:)) < 10
filenames_numbers(i,:) = ['0' filenames_numbers(i,2)];
end
end

filenames = cellstr(num2str(zeros(96,1)));
wellnames = filenames;
c = ['_', group ,'_SAPA-23.csv'];
cID = ['_', group ,'_SAPA-23'];

for i = 1:length(filenames)
plateID(i) = {[filenames_letters(i) filenames_numbers(i,:)]}';
filenames(i) = {[[filenames_letters(i) filenames_numbers(i,:)], c]};
wellnames(i) = {[[filenames_letters(i) filenames_numbers(i,:)], cID]};
end

% Rip sequences
seq_long = cell(96,1);
foldername = [group, '_sapa23_csv\'];
for i = 1:96
[num, txt] = xlsread([foldername, filenames{i}], 'A2:A2');
seq_long(i) = txt;
end

% Define sequences
seq_plasmid = [[See other appendices]];
aa_plasmid = DNAtranslator(seq_plasmid(1:end-2));
aa_wt = [[See other appendices]];
```

```
seq_wt = seq_plasmid(668*3-2 : 668*3-2+117*3-1);
aa_comp = [[See other appendices]];
seq_comp = seq_plasmid(689*3-2 : 689*3-2+length(aa_comp)*3-1);

preZ = seq_plasmid(2064-19:2064);
postZ = seq_plasmid(2065+length(seq_comp):2065+length(seq_comp)+19);

% Trim and classify
seq = seq_long;
comment = cell(96,1);
error_fail = 'Sequencing failed';
error_short = 'Sequence too short to align successfully';
error_N = 'Z-domain contains undetermined bases. Check reverse sequence';
error_indel = 'Sequence contains insertions or deletions';
error_wt = 'Sequence is wild-type';
error_align = 'Alignment failed';

for i = 1:96
% Cut-out comp. seq.
if length(seq_long{i}) >= 174+2*20
Zstart = strfind(seq_long{i}, preZ) + length(preZ);
Zend = strfind(seq_long{i}, postZ) -1;
seq{i} = seq_long{i}(Zstart:Zend);
elseif length(seq_long{i}) == 5 && mean(seq_long{i} == 'NNNNN') == 1
comment{i} = error_fail;
seq{i} = seq_long{i};
elseif isempty(comment{i}) == 1 && length(seq_long{i}) < 174+2*20
comment{i} = error_short;
seq{i} = seq_long{i};
else
comment{i} = error_align;
seq{i} = seq_long{i};
end
% Is comp. seq. longer/shorter/wt?
if isempty(comment{i}) == 1 && length(seq{i}) ~= length(seq_comp)
comment{i} = error_indel;
elseif isempty(comment{i}) == 1 && length(seq{i}) == length(seq_comp)
if seq{i} == seq_comp
comment{i} = error_wt;
end
end
% Does comp. seq. contain bases other than ACTG?
if isempty(comment{i}) == 1 && length(seq_long{i}) >= 174 ...
&& length(strfind(seq{i}, 'A')) + length(strfind(seq{i}, 'T')) ...
+ length(strfind(seq{i}, 'C')) + length(strfind(seq{i}, 'G')) < length(seq{i})
comment{i} = error_N;
elseif isempty(comment{i}) ~= 1 && length(seq_long{i}) >= 174 && ...
length(strfind(seq{i}, 'A')) + length(strfind(seq{i}, 'T')) ...
+ length(strfind(seq{i}, 'C')) + length(strfind(seq{i}, 'G')) < length(seq{i})
comment{i} = [comment{i},' & ', error_N];
end
end

% Translate sequences
aa = cell(96,1);
counter = 0;
for i = 1:96
counter = counter +1;
```

```
if isempty(comment{i}) == 1
aa{i} = DNAtranslator(seq{i});
else
aa{i} = '-';
end
end

% Compare to wt seq
subsIndex = cell(96,1);
subsFrom = subsIndex;
subsTo = subsIndex;

for i = 1:96
if isempty(comment{i}) == 1
subsIndex{i} = find((aa{i} == aa_comp) == 0);
subsFrom(i,1) = {aa_comp(subsIndex{i,1})};
subsTo(i,1) = {aa{i}(subsIndex{i,1})};
else
subsFrom(i,1) = {'-'};
subsTo(i,1) = {'-'};
subsIndex{i,1} = '-';
end
end

% Switch index to position
position = cell(96,1);
for i = 1:96
if isnumeric(subsIndex{i}) == 1
position{i} = subsIndex{i} + 21;
else
position{i} = subsIndex{i};
end

if length(position{i}) > 1
position_merge = zeros(1,length(position{i}));
for j = 1:length(position{i})
position_merge(j) = position{i}(j);
end
position{i} = num2str(position_merge);
comment{i} = 'Multiple substitutions';
end
end

% Sort

for i = 1:96
if isnumeric(position{i}) == 1
position_sort(i,1) = position{i};
else
position_sort(i,1) = 0;
end
end

[sortVal, sortIndex] = sort(position_sort);

output = cell(96,8);
output(:,1) = wellnames(:,1);
output(:,2) = comment(:,1);
```

```
output(:,3) = seq_long(:,1);
output(:,4) = seq(:,1);
output(:,5) = aa(:,1);
output(:,6) = position(:,1);
output(:,7) = subsFrom(:,1);
output(:,8) = subsTo(:,1);

output_sorted = cell(96,8);

for i = 1:96
for j = 1:8
output_sorted{i,j} = output{sortIndex(i),j};
end
end

headers = {'Well name', 'Comment', 'Full NT sequence' ...
'Aligned NT sequence', 'Aligned AA sequence', 'AA position'...
'From AA', 'To AA'};
output_sorted

xlswrite('results.xls', headers, group, 'A1:H1')
xlswrite('results.xls', output_sorted, group, 'A2:H97')

end
```

F aa substitutions

Table 6: High binders at IgG concentration 0.4 nm

Wells 1-48

Well name	AA position	From AA	To AA
A01	-	-	-
A07	-	-	-
C09	-	-	-
D04	-	-	-
D08	-	-	-
D09	-	-	-
D12	-	-	-
H11	-	-	-
F05	-	-	-
E04	-	-	-
H02	-	-	-
G07	22	V	S
B10	26	F	H
G08	26	F	I
D06	26	F	N
G02	26	F	N
G11	26	F	P
C12	26	F	S
D11	26	F	S
A08	27	N	A
F01	27	N	A
G03	27	N	A
A09	27	N	E
A10	27	N	E
A12	27	N	E
F06	27	N	E
A06	27	N	F
G12	27	N	G
B03	27	N	H
B05	27	N	H
C01	27	N	H
C10	27	N	I
E02	27	N	I
E03	27	N	I
F04	27	N	I
H04	27	N	I
H10	27	N	K
C08	27	N	L
E09	27	N	L
H12	27	N	L
A11	27	N	M
H06	27	N	M
A03	27	N	Q
B06	27	N	Q
C04	27	N	Q
C05	27	N	Q
C06	27	N	Q
D03	27	N	Q

Wells 49-96

Well name	AA position	From AA	To AA
D07	27	N	Q
D10	27	N	Q
G10	27	N	Q
A04	27	N	R
C02	27	N	R
C07	27	N	R
D01	27	N	R
E07	27	N	R
E08	27	N	R
B07	27	N	V
C11	27	N	V
F02	27	N	V
B08	27	N	W
F09	27	N	W
F12	27	N	W
F10	27	N	Y
F03	28	K	R
E10	29	E	F
G06	29	E	G
B09	29	E	I
B02	29	E	L
E11	29	E	L
A02	29	E	V
H08	35	Y	W
E05	36	E	D
B04	41	P	A
F11	41	P	G
D02	43	L	V
F07	44	N	L
G01	44	N	L
G05	44	N	L
C03	46	E	Q
E12	46	E	Q
H07	47	Q	G
G04	54	S	H
G09	58	D	A
H09	58	D	A
D05	59	P	E
A05	61	Q	S
H01	65	L	E
B11	65	L	M
B12	65	L	N
F08	65	L	S
E06	67	A	L
B01	72	L	I
H03	75	A	H
H05	75	A	H
E01	75	A	P

Table 7: High binders at IgG concentration 10 nm

Wells 1-48

Well name	AA position	From AA	To AA
F12	-	-	-
E07	-	-	-
A09	27 79	NK	KH
C02	42 64	NN	YD
D11	44 79	NK	TQ
E02	54 79	SK	VW
G10	62 71	SK	RV
F05	22	V	P
F07	23	D	M
A08	23	D	Y
B11	24	N	D
E10	24	N	F
B07	24	N	Q
E03	24	N	Q
B09	25	K	C
A02	25	K	M
H02	25	K	N
C03	26	F	M
A11	26	F	S
F01	26	F	S
F08	27	N	D
D03	27	N	G
B03	27	N	P
G02	28	K	T
B12	29	E	I
E09	29	E	P
C05	30	Q	K
C07	32	N	I
G11	32	N	R
D12	36	E	K
D01	36	E	M
D07	36	E	V
C11	36	E	W
G01	38	L	A
F04	39	H	M
G07	41	P	M
C08	42	N	W
D05	44	N	L
H06	45	E	A
D08	45	E	M
F09	45	E	T
H07	46	E	F
A06	47	Q	E
E06	47	Q	E
A03	48	R	A
F11	48	R	H
G04	48	R	I
G03	48	R	W

Wells 49-96

Well name	AA position	From AA	To AA
C10	49	N	D
E01	49	N	Q
F10	49	N	Q
H09	49	N	R
H12	50	A	E
C04	51	F	K
A12	51	F	S
C09	51	F	Y
H10	52	I	F
G05	52	I	L
G09	53	Q	E
H05	54	S	V
E12	56	K	L
H03	56	K	L
F06	57	D	E
D09	57	D	F
B04	57	D	V
B08	58	D	E
H01	60	S	G
E05	60	S	N
F02	61	Q	D
E08	63	A	G
A07	63	A	T
H11	64	N	K
A04	65	L	R
A05	65	L	S
B05	65	L	T
B02	65	L	Y
E04	67	A	E
F03	67	A	I
G06	67	A	P
H04	67	A	Q
D10	67	A	R
C06	67	A	T
D02	68	E	D
C01	71	K	M
H08	72	L	M
B10	75	A	K
D06	76	Q	F
A10	76	Q	K
E11	76	Q	T
B01	77	A	I
B06	78	P	D
D04	78	P	F
G08	78	P	R
A01	78	P	W
G12	79	K	C
C12	79	K	V

Table 8: Medium binders at IgG concentration 10 nm

Wells 1-48

Well name	AA position	From AA	To AA
F09	-	-	-
G01	-	-	-
H06	-	-	-
B10	-	-	-
H09	-	-	-
C03	-	-	-
D01	-	-	-
C02	30 33	QA	WS
D05	41 53	PQ	GP
E02	62 71	SK	RV
A12	62 71	SK	RQ
B07	57 62	DS	HI
F04	22	V	C
A10	24	N	C
A06	25	K	F
G09	26	F	K
H08	26	F	L
H12	30	Q	D
B01	30	Q	E
G03	30	Q	G
B03	30	Q	I
E05	30	Q	I
E01	30	Q	K
A01	30	Q	L
A08	30	Q	P
E08	30	Q	T
D11	30	Q	Y
G04	32	N	C
F06	32	N	P
A02	34	F	Y
E07	35	Y	H
H11	35	Y	T
F11	36	E	K
H05	38	L	A
F12	38	L	R
C08	39	H	F
H03	39	H	G
F08	39	H	L
H01	39	H	L
A11	39	H	M
E09	39	H	M
C07	39	H	P
H07	39	H	P
C11	39	H	S
D02	39	H	S
G08	39	H	S
G11	39	H	S
H10	39	H	S

Wells 49-96

Well name	AA position	From AA	To AA
E04	39	H	T
G07	39	H	T
G10	39	H	V
D09	43	L	C
G05	43	L	C
A07	43	L	D
D03	43	L	E
F03	43	L	G
H04	43	L	G
D10	43	L	K
F07	43	L	K
B12	43	L	N
C06	43	L	N
B09	43	L	P
C09	43	L	Q
G06	43	L	Q
B11	43	L	S
B02	45	E	L
G02	45	E	R
C04	45	E	W
B08	47	Q	K
A09	48	R	P
C10	49	N	H
E12	49	N	L
E10	49	N	M
E06	49	N	S
F10	49	N	V
A05	49	N	W
D12	51	F	G
C05	52	I	F
A03	52	I	H
E03	52	I	L
B05	52	I	R
E11	53	Q	E
G12	53	Q	L
B06	53	Q	P
F02	53	Q	P
C12	53	Q	W
A04	56	K	F
F01	56	K	S
C01	65	L	D
F05	65	L	G
D07	67	A	M
D04	72	L	C
H02	75	A	H
D08	78	P	I
B04	79	K	M
D06	79	K	V

Table 9: Low binders at IgG concentration 10 nm

Wells 1-48

Well name	AA position	From AA	To AA
B04	-	-	-
F03	-	-	-
H04	-	-	-
C03	35 45	YE	HG
C11	35 45	YE	HG
D02	38 75	LA	WG
F01	35 50	YA	SR
H09	56 75	KA	YC
H12	31 52	QI	RN
F06	26	F	A
C02	31	Q	C
D08	31	Q	E
D12	31	Q	E
H03	31	Q	E
H06	31	Q	E
A02	31	Q	F
D06	31	Q	F
E11	31	Q	F
D10	31	Q	G
F12	31	Q	G
G02	31	Q	H
A03	31	Q	L
G05	31	Q	L
A08	31	Q	R
D07	31	Q	T
E03	31	Q	T
G12	31	Q	T
B05	31	Q	V
D11	31	Q	V
A12	31	Q	W
H08	31	Q	W
C08	31	Q	Y
E12	32	N	P
F10	34	F	C
F08	34	F	D
F09	34	F	D
A06	34	F	E
G06	34	F	E
E01	34	F	H
F04	34	F	H
E09	34	F	I
G04	34	F	N
H02	34	F	P
H11	34	F	P
A09	34	F	Q
E10	34	F	Q
B10	34	F	S
D03	34	F	S

Wells 49-96

Well name	AA position	From AA	To AA
A04	34	F	T
E02	34	F	T
D04	35	Y	E
H05	35	Y	E
H01	35	Y	G
B11	35	Y	H
A10	35	Y	K
B09	35	Y	L
C04	35	Y	M
C05	35	Y	M
B08	35	Y	N
C10	35	Y	N
B06	35	Y	P
E07	35	Y	P
F02	35	Y	P
D05	35	Y	S
B01	38	L	C
E06	38	L	E
G09	38	L	E
C06	38	L	N
B07	38	L	P
H07	38	L	W
D09	39	H	A
A07	39	H	D
C12	39	H	D
E08	39	H	D
F05	39	H	D
G08	39	H	E
C01	39	H	S
E04	43	L	D
A05	43	L	E
B02	52	I	D
F11	52	I	D
G11	52	I	D
C07	52	I	E
F07	52	I	G
B12	52	I	N
C09	52	I	P
G07	52	I	R
G01	52	I	S
G03	52	I	S
A01	52	I	W
D01	52	I	W
E05	52	I	W
B03	56	K	E
A11	56	K	P
G10	56	K	P
H10	56	K	P