Bachelor Degree Project

UNIVERSITY OF SKÖVDE

CLASS SWITCHING MECHANISM IN ANTIBODIES

A conformational and structural investigation of the B-cell repertoire diversity of the CDR3

Version 2
01/2021-06/2021

Bachelor Thesis Project in Biomedicine

30 ECTS, Spring term 2021

Sofia Fredin

A17soffr@student.his.se

Supervisor: Franca Fraternali

Franca.fraternali@gmail.com

Examiner: Katarina Ejeskär

Katarina.ejeskar@his.se

School of Health Sciences

University of Skövde

Box 408

541 28 Skövde

Abstract

A process called class switching recombination (CSR) regulates the production of isotypes by enabling the switch of producing unspecific IgM or IgD to producing IgG, IgE, or IgA by causing an irreversible genomic rearrangement. Recent studies indicate a gap in the identification of the molecular mechanisms underlying CSR. This interdisciplinary project between bioinformatics and immunology aims to shed light onto these mechanisms and highlight differences in physicochemical properties in the B-cell repertoire to identify factors contributing to the decline in immune response with age and its relationships to CSR. The project is developed in line with the hypothesis that effective immune responses, and their resolution, require coordinated action from multiple different subclasses of antibodies. This is enabled by improvements in the high throughput sequencing techniques, which allows the B-cell repertoire to be revealed. The most variable portion of antibody molecules is the third complementarity determining region (CDR3) of the heavy chain. Hence, differences in hydrophobicity, pK value of the C-terminus peptide fragment, and the length of CDR3 were identified and accounted as factors that contribute to the decline in the immune response with age. Our understanding of how the immune response alters with age, is important for research applications in designing tailored vaccine strategies to ensure safety and efficacy. Development of *in silico* tools for the analysis of the B-cell repertoire is useful in extracting principles and functioning mechanisms that can be tested experimentally, helping in further characterizing CSR.

Popular Scientific Summary

Our immune system is crucial for maintaining health and providing the body with protection against pathogens. Parts of some functions within the immune system have been shown to decline with age, and this report aims to account for the identified differences in the immune system between young and old individuals. By understanding how the immune system works and alters with age, this basic research could be applied to the improvement of vaccine strategy and regimen in order to ensure the safety and efficacy of vaccines. Differences in functions of the immune system were compared by using bioinformatic tools. Bioinformatics is an interdisciplinary field where computational biology is combined with molecular biology, which enables large sets of data that consist of highly valued information about the molecular mechanisms to be stored and applied to an object of interest. The B-cell repertoires was collected from in total 6 different donors of the ages 24, 26, 34, 55, 56 and 59 at the time points day 0, 7, 14 and 28 post vaccination. The samples were divided into age groups of young (>34 years) and old (<55 years) to compare alterations in immune response with age.

Our immune system consists of multiple different cell types, that all cooperate to eliminate pathogens and clear infections. One of the cell types that are responsible for generating a memory to certain pathogens are B-cells, which will secrete antibodies that will bind to the pathogen and neutralize it. The DNA sequence of the antibodies collected from the individuals was collected via a laboratory technique and deposited into a database. The extracted sequences will contain highly valued information about how the individual has responded to immune system stimulation.

The collected sequences were then compared in order to find differences in their physicochemical properties, aiming to give an insight into how the structure of the antibody will alter with age. The analysis performed in BRepertoire indicated that the antibodies produced by older people tended to have a longer, more hydrophobic fragment at the site that regulates the binding to pathogens. These factors cause alterations in how the antibodies can bind to their target, which has a negative impact on their ability to clear the infections and thus cause an impaired ability to generate an efficient immune response.

Investigating the structure of antibodies has several different implications for science. The use of antibodies as therapeutics in the drug discovery and development industry is expanding and is showing promising effects in clinical trials. Collecting data and knowledge regarding the structure and function of the DNA sequences of the antibodies can contribute to a fundamental discovery in basic research. The research can also point out differences in the regulators of the immune system that can be used as biomarkers for a dysfunctional immune system in the future. Using biomarkers could potentially increase the accuracy of diagnostic procedures and reduce the time frame from showing symptoms to receiving a proper diagnosis. This would enable the patients to receive the correct diagnosis, in time, to get the right treatment and increase the life quality and survival rate of the patients.

The application of bioinformatics is multifunctional and can be used in the drug discovery and development process which will decrease the requirements of animal models. The results generated with bioinformatic tools might even provide a more accurate analysis than the results obtained from animal models, since the outcome from the animal model may not always have the same mechanism of action on human cells. The application of bioinformatics also allows the management of big sets of data that holds extensive amounts of information that can be applied to both healthcare and drug discovery and development processes.

Abbreviations

AID	Action-induced cytidine deaminase
CDR3	Complementary Determining region 3
DBS	Double Strand Breakage
Fab region	Fragment antigen-binding
Fc region	Fragment crystallizable region
In silico	Computer-based experiments
In vivo	Experiments inside living organisms
In vitro	Experiments on livning organisms outside their normal environment
NHEJ	Non-homologous end-joining
VTD	Viscerotropic disease
NGS	Next Generation Sequencing

Table of Contents

Introduction……1
Material & Methods……5
Results……8
Discussion……13
Acknowledgements……19
References……20

Introduction

The physiological mechanisms protecting the human cells against an invasion of pathogens can be observed by studying the immune system. An operative immune system is crucial for the survival and maintenance of human health, and furthermore, dysfunctions or deviations could potentially be fatal. The immune system is diverse and acts via multiple mechanisms to prevent a pathogen from replicating within our cells and potentially cause disease (Berthelot et al., 2013). The immune system is divided into the innate and the adaptive immune system depending on the mechanism of action and will cooperate in eliminating pathogens (Iwasaki & Medzhitov, 2015).

The innate immune system provides the first line of defense and includes both chemical and physical barriers such as skin, mucosa, saliva, and sweat. In addition to the barriers, the innate immune system consists of several immune cells acting to neutralize the antigen, including cytokines, macrophages, monocytes, and neutrophils (Parkin & Cohen, 2001). The innate immune system operates fast and serves as the first line of defense while the adaptive immune response contributes to memory cells required for achieving immunity to a disease (Vipula & Atula, 2018).

The adaptive immune responses are tailored to specific pathogens and developed over time (Iwasaki & Medzhitov, 2015). T-cells and B-cells are important components of the adaptive immune response. T-cells are responsible for inducing a cell-mediated immune response and the B cells will act to produce plasma cells and memory cells (Vipula & Atula, 2018). B-cells will produce different subclasses of antibodies, aiming to improve the recognition of an antigen. The diversity in the B-cell repertoire is crucial for an effective immune response. Elderly people have been showed to have a decline in diversity, which is one of the suggested reasons for the decline in the immune response. A suppressed immune system results in a decreased protection to pathogens (Parkin & Cohen, 2001).

Antibodies are made up of glycoproteins and have a y-shaped structure, consisting of in total two light chains and two heavy chains (Figure 1). The heavy chain is in its turn further divided into two parts, a constant region and a variable region of the antibody. The variable region of the antibody is called the Fab region and will adapt in order to attach to antigens, and the constant region of the antibody is called the Fc region and is responsible for the activation of other cellular mechanisms involved in the immune defense. The constant region will also define the type of antibody (Tanabe & Kajihara, 2020). The constant region of the antibody will change according to a process called class switch recombination (CSR), which will result in an irreversible arrangement of genes (Stavnezer et al., 2008). B-cells will proliferate and differentiate into different subclasses in two different ways, depending on the mechanism diversifying the B-cell repertoire, namely Class switch recombination (CSR) or somatic hypermutation. This report will have a focused approach to the mechanism behind CSR. The CSR will enable the B-cells to switch from producing unspecific isotypes IgM and IgD to produce specific isotypes: IgA, IgE, and IgG (Chi, Li & Qiu, 2020). The different isotypes include IgG (1-4), IgD, IgA (1,2), IgM, and IgE. Naïve B-cells will switch isotype from unspecific IgM and IgD after activation by an antigen

to differentiate into one of the specific isotypes, IgG, IgA, or IgE (Stavnezer et al., 2008).

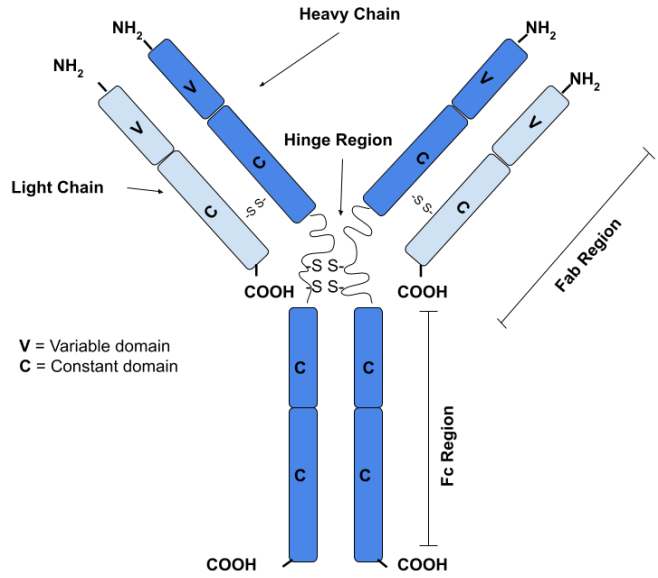

Figure 1. Illustration of the structure of an antibody. The light chain of the antibody is expressed as the light blue colored two chains of the upper antibody, and the heavy chain is colored darker blue. The upper part of the antibody marked with "V" illustrates the variable region of the antibody while the other regions marked with "C" are parts of the constant region.

The heavy chain is encoded by chromosome 14 in humans. More specifically, the variable region of the heavy chain is encoded by different segments, including variable (V), diversity (D), and joining (J), which make up for the VDJ segments. The variable region of the light chain is encoded by exclusively V and J. Class switch recombination cannot occur without a functional arrangement of segments since the synthesis would be inhibited, which would result in apoptosis of the cells (Chi, Li & Qiu, 2020). The downstream site of the region encoding for the VDJ segments contains Cμ, which causes an expression of IgM on the cell surface acting as a receptor (Tanabe & Kajihara, 2020).

The variable region of the heavy chain consists of complementary determining regions (CDR), built up by short peptide fragments, responsible for the binding to antigens (Figueiredo et al., 2014). There are three different CDRs present in the variable region, CDR1, CDR2, and CDR3. CDR3 has been shown to be the most variable and is therefore playing a crucial role in the antigen-binding since it modulates the joining of the different gene segments of the region. Alterations in the CDR3 structure impact the rearrangement of the genes, which adapts the variable regions' affinity for antigens (Townsend et al., 2016).

The heavy chain has an upstream structure including motifs of conserved nucleotides called switch region (S-region) (Zhang et al., 2019). The double-stranded DNA located within the S-regions in the gene segment will be split up by an enzyme called action-induced cytidine deaminase (AID). AID will initiate the single-strand breakage on both strands, which results in a double-strand breakage. AID regulates the mRNA sequence in order to act on free deoxycytidine present in the cell (Bransteitter et al., 2003). The process causes the single-strand DNA located on the S region to break, which requires it to be further broken down into double-strand breaks to enable the structural changes in DNA (Stavnezer et al., 2008). Translocating components in the gene segments enables the alteration in structure encoding the genes in the heavy chain of the antibody. The exons located in between the S regions will be removed from the gene segment when the DNA segments are ligated back together (Figure 2) via non-homologous end joining (NHEJ) (Lieber, 2010). The DBS in the DNA segments will be rejoined by NHEJ, resulting in recombination in the gene segment, enabling CSR.

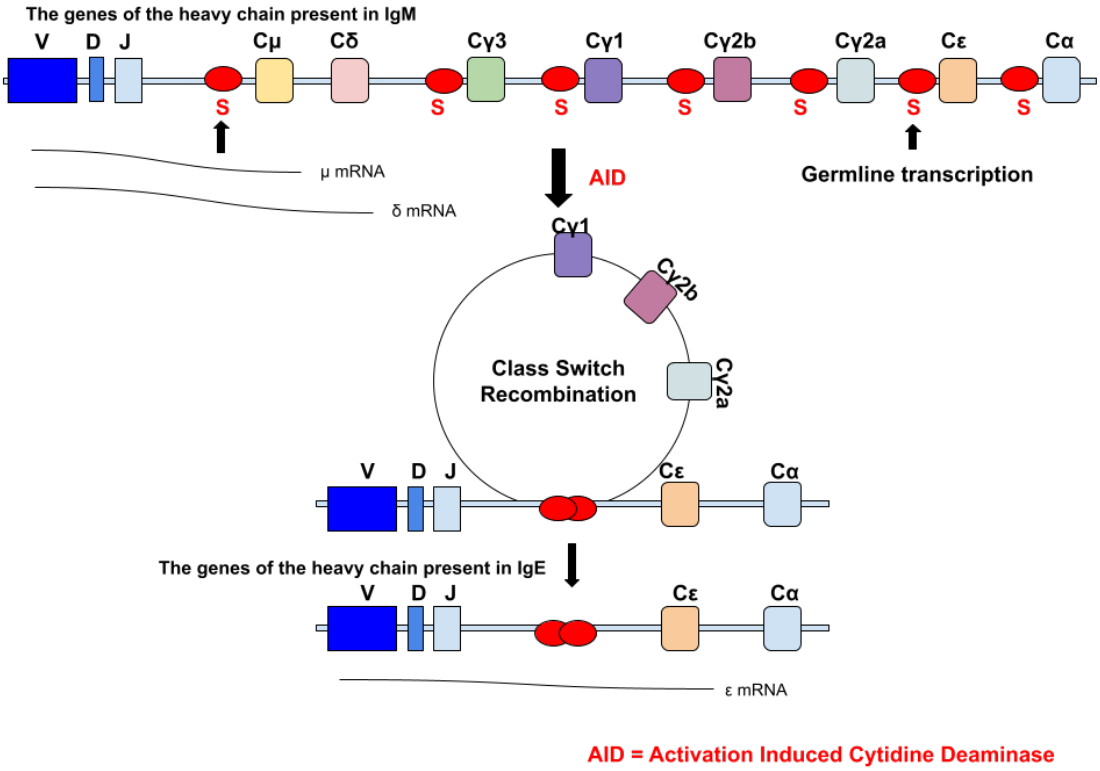

Figure 2. An overview of the class switch recombination process. The gene segments present in IgM undergo class switching recombination to get transcribed into IgE. The top segment expresses the genes present in an IgM antibody, illustrating that CSR is enabled by AID between S regions. The gene segments in between the S regions are transcribed and the germline transcription occurs. The gene segments second from the top line express the mechanism behind the deletion of regions in the gene segments occurring in between two S regions. The third line shows the rearranged gene segment.

Immunization against a disease can be achieved by vaccinations. A vaccine will introduce an antigen to the immune system without causing disease, aiming to contribute to the recognition of the antigen in a secondary infection. Vaccinations against severe diseases were introduced in Western Europe in 1796 by Edward Jenner, which eventually led to the eradication of the infectious disease smallpox (Parham, 2015). Immunization is dependent on the variable region's ability to adapt to increase its affinity to antigens, contributing to an efficient immune response. The switching of classes is regulated by CSR and will cause the naive B-cells to stop producing unspecific IgM and IgD and start producing specific subclasses. The mechanism of action behind vaccination is highly dependent on CSR since it will direct the immunoglobulins produced which are responsible for the protection in a secondary infection of the antigen (Frasca et al., 2016). Elderly patients have been shown to have a decline in immune response, which limits the protective effect of the vaccines (Dunn-Walters & Ademokun, 2010).

Recently improved techniques of high-throughput sequencing of B-cells enable the repertoire of the B-cells to be examined. During this experiment, B-cell repertoires collected from patient samples will be investigated. The B-cells originate from patients who received the 17D Yellow Fever Vaccine, which is an attenuated viral strain vaccine responsible for triggering a coordinated immune response including activation of both innate and adaptive immune response (Davis et al., 2019). Understanding how the interplay of the variable and constant regions of the heavy chain might cause differences in structure and function of the antibodies would facilitate and increase the efficacy of the drug discovery and development of the field. The application of bioinformatics software effectively integrates with the laboratory work and facilitates managing big sets of data, generating more reliable results.

Aim and hypothesis

This is an interdisciplinary project between immunology and bioinformatics. The report will have a focused approach on the in-silico tasks performed, in order to investigate B-cell repertoires. The project aims to identify endogenous and exogenous factors involved in the mechanism of CSR, aiming to develop a wider understanding of the molecular mechanism. In addition, the repertoire of the B-cells originating from different donors will be compared in order to identify differences between old and young age groups. The structural and conformational investigation of the repertoires will reveal differences in physicochemical properties contributing to the decline in immune response with age. A full identification of the mechanism behind CSR and how the adaptive immune response is altered with age are central for an improvement of the vaccination regimen and the development of biological drugs since a complete identification of the mechanism behind CSR would enable an extensive field of drug discovery and development to expand. The project develops in line with the hypothesis that effective immune responses, and their resolution, require coordinated action from multiple different subclasses of antibodies.

Materials & Methods

The B-cells were sequenced to reveal the structure of the antibody amino acid sequence. This enabled the different chains of the antibody to be defined and exposed the structure and properties of the antibody's amino acid sequences. This was performed by sequencing the B-cells in next-generation sequencing (NGS). NGS enabled the sequencing of extensive amounts of genomic data within a short period of time, which makes the technique both cost and time-effective (Reis-Filho, 2009). The high-throughput NGS used during this *in silico* based experiment revealed the sequences of the antibodies present in the donor samples, which provided highly valued information about the characteristics and properties of the antibodies (Martin et al., 2016).

B cell scRNA sequencing will be performed in order to explore the repertoire of the B-cells and determine their expression profiles. Isolating single cells enables the VDJ region structure to be examined, which collects highly valued information about the antibody structure. This enabled the identification of subclasses present in the samples, which contributed to the identification of the immune response to an antigen. The information generated by examining the repertoire of the B-cells from scRNA sequencing will express critical information about the molecular mechanisms that control the class switching of the antibodies. The cells included in the dataset were prepared and sequenced with high-throughput sequencing and analyzed with the software BRepertoire (Margreitter et al., 2018).

This report is based on a set of data generated from six healthy volunteers receiving the Yellow Fever 17D Vaccine. The B-cell repertoires was collected from donors of the ages 24, 26, 34, 55, 56 and 59 at the time points day 0, 7, 14 and 28 post vaccination. The samples of the donors were divided into two different age groups of young (>34) and old (<55). In total, 33 000 sequences of the heavy chain were identified and sequenced for the structural and conformational investigation.

The dataset exclusively consists of sequences of the heavy chain, concluding the analyses to observe the CDR3 region of the isolated heavy chain. The extensive impact on the immune response of differences in the CDR3 properties implies an interest in using CDR3 as a biomarker for future perspectives. The physicochemical properties affect the whole protein folding and structure beyond the binding affinity to different molecules. The Kidera Factors expressed in Table 1, were investigated in order to identify differences in physicochemical properties of the CDR3 between age groups (Dunn-Walters et al., 2018).

The dataset based on samples generated via next-generation sequencing was aligned in IMGT/HighV-QUEST before exported into BRepertoire in order to analyze the rearranged sequences of nucleotides in the samples. This software was able to analyze the nucleotide arrangement origin from the immunoglobulins present in the samples, which enables the adaptive immune response to be examined in an accurate way. It acts by aligning the nucleotide sequences with a reference gene in order to enable the identification of the different gene segments (Alamyar et al., 2012).

BRepertoire is a webserver implemented in R in the groups of Franca Fraternali and Deborah Dunn-Walters in the context of the MABRA consortium. This tool facilitates visualizing repertoires of B-cells and computes properties and structures of sequences (Margreitter et al., 2018). R is implemented in BRepertoire and applies statistical methods as indicators of significance at a threshold of significance set to >0.05.

In this investigation, BRepertoire was used in order to detect differences in the physicochemical properties of the antibodies generated by the different age groups. The investigated factors included the Kidera Factors, which are calculations of the physicochemical properties of the structures that

predicts how they will behave in certain environments (Epstein et al., 2014). The length of the CDR3 amino acid sequence was investigated in order to draw conclusions regarding the flexibility of the antibody structure, which will impact the interaction characteristics. A constricted flexibility of the antibodies has been associated with an enhanced affinity for the antigens. The length of the CDR3 amino acid sequence will affect the flexibility of the variable region, which will have an impact on the interactions with immunoglobulins (Li et al., 2015). In addition to the CDR3 amino acid sequence length, other factors such as subclass frequency and gene frequency were investigated. The comparison of the frequency of subclasses and genes between different age groups provides an overview of the difference in immunoglobulin production with age. The gene frequency was investigated by analyzing the composition of the VDJ segments in BRepertoire. The factors age, "Vfamily", "Dfamily", and "Jfamily" were grouped to generate a diagram of frequency comparing the different gene segments (Margreitter et al., 2018).

Firstly, the dataset was uploaded at "IMGT/HighV_Quest" in order to align the nucleotide sequences onto a reference genome. Next, the dataset was uploaded to BRepertoire into the branch "IMGT/V_QUEST management" "upload" tab in BRepertoire in order to make the analysis of the nucleotide sequences even more extensive. This branch calculates meta-data of the sequences and generated a calculation of the VDJ segment matches. The output of the analysis contained columns with the content of the calculated branches. This output table was downloaded and uploaded into the "Annotation" branch of the "IMGT/V_QUEST management". This allowed all results generated in the output file from the "upload" to be expressed as one spreadsheet. This was used to ensure the informative content was the expected data. The calculation branch will implement additional analysis beyond the ones provided by the "IMGT/V_QUEST management" branch.

The output from "upload" at the "IMGT/V_QUEST management" branch was uploaded to the calculation branch at the "upload" tab. The separator used for the data file was set to "comma" and the quote for strings "double quote". The "calculate" tab in the property calculation branch was used in order to calculate the physicochemical properties of the sequences in the dataset. "CDR3.IMGT" was selected as an input column for calculation. The output was downloaded as a CSV file to use in later analysis. The data was uploaded onto the "extract" tab at the "property calculations" branch and "V.GENE.and.allele" was chosen as the specific input column in order to enable a more detailed analysis of the V genes. "String indices" was chosen as the extraction method and the range of sub-string 8-12 and the new column was extracted and named "V-gene".

The output from the calculation of the physicochemical properties from the calculation branch was uploaded onto the "data analysis" branch. The comma (,) was used as a separator and a double quote was used as quotes for strings. From this step, the columns of interest were selected in the "select" tab. The data columns selected was "class", "subclass", "Vfamily", "JFamily", "Dfamily", "V.GENE.and.allele", "CDR3.IMGT", "AA_CDR3_edited", "Donor", "Day", "Age", "CDR_aa_seq_edited", "CDR_aa_length" and the 10 Kidera Factors listed in Table 1.

Table 1. The analyzed Kidera Factors and their measured physicochemical property.

Kidera Factor	**Physicochemical Property**
1	Helix Bend
2	Side Chain Size
3	Extended Structure Preference
4	Hydrophobicity
5	Double bend
6	Partial Specific Volume
7	Flat Extended Preference
8	Occurance Alpha Reg
9	pK.C

The “filer” tab was used to exclude partial data from the data analyzed. “IgL” was excluded from the raw data since this was included in the original dataset. The named parameters “day” “age” and “donor” were also filtered out at this step. “Grouping” was performed to enable the dataset to be compared and put the age groups against each other. Subclass, donor, age, and day were grouped together during step 1 select grouping columns. The plots and graphs of the data were generated with “select analyses” and the statistical tests were added.

A box-whisker plot was performed in order to express the differences in amino acid sequence length of the CDR3 since it’s an appropriate plot for representing numerical data. Wilcoxon rank-sum test was used as a statistical test, with a significance level of >0.05. The data was split into “age”, “day” and “donor” and the data was grouped for donor and age, which enabled comparisons of the CDR3 amino acid sequence length between different days and patient groups. A bar plot of the subclass frequency was generated in order to investigate the differences in levels of immunoglobulins present in the samples. Once again, the donor and age were paired in data grouping, and the relative values in percentage were normalized by all data present in the sample. The Kidera Factors were investigated by generating a distribution analysis. This provides an overview of the distribution of data, and the physicochemical properties of the immunoglobulins could be examined and compared. R was implemented and calculated the statistical differences with a significance threshold below 0.05. Cliff’s delta was chosen as an effect size plot with a confidence interval of 90%.

Antibody modeling was performed in ABodyBuilder to examine the antibody repertoire and predict the side chain of the antibody ABodyBuilder (Dunbar et al., 2016). The confidence threshold was set to 75%. Antibody Visualization was performed using PyMol, by exporting the paired sequences of the heavy and light chain from ABodyBuilder. In this way, visualization of two antibody structures can be enabled in the same figure, allowing a visual comparison of the structure between an antibody structure generated in an old and a young donor in PyMol (The PyMol Molecular Graphics System, Version 2.0 Schrödinger, LLC). The antibody structure visualization was based on the sequences and the data expressed in Table 2.

Table 2. Amino acid sequences of light and heavy chain, age, and donor ID of the data used to build the model of the antibody structures by pairing the chains in ABodyBuilder and visualizing the aligned models in PyMol.

Donor ID	**189**	**200**
Age	59	24
Amino acid sequence of heavy chain	DVQLVESGGGLVQPG GSLTLSCVASGFTFRS YWMQWVRQVPGKG LAWVSRINNDGSGTT YADSVRGRFTISRDN AKNTLYLQMNSLTA EDTAIYYCGRDQDGR GGTIDYWGQGTRVTVSS	AATSTSRGAGLLKPSE TLSLICAVYGGSFTDS YWTWIRQPPGKGL EWIGKINHSGSTKY NPSLKSRITLSVDTS KNQFSLKLSSVTAAD TAVYYCAREMERPTV TTFWYDSWGQGTLLT VSS
Amino acid sequence of light chain	DIQMTQSPSSLSASVGDRVTITC RASQSISSYLNWYQQKPGKAP KLLIYAASSLESGVPSRFSGSG SGTDFTLTISSLQPEDFATYYC QETYITPLTFGGGTKVEIK	AYELTQPPAVSVSPG QTARITCSGDAFPRQ YSYWYQKRPGQAPV LVIYKDNERPSGIPE RFSGSTSGATVTLIIS EVQAEDEADYYCQS ADNSGAYHFFGGGT KLTVL

Results

The results generated in BRepertoire provide an overview of the differences in physicochemical properties of antibodies originating from donors of different ages. In addition to the physicochemical properties, the frequency of different subclasses was investigated.

A bar plot of the subclass frequency was generated in BRepertoire in order to examine the differences in immune response between the age groups (Figure 3). As Figure 3 suggests, the predominant response in both age groups was a secretion of IgM followed by IgA1 and IgG1. The examined subclasses only include IgA, IgG, and IgM, which causes the generated results to exclusively be applicable to the immune response involving the specific isotopes. The secretion of IgE is not expected as a response to a vaccine unless the donor is allergic (Chung, 2014), and IgD coexists with IgM predominantly (Geisberger et al., 2006).

Barplot of subclass frequency between the age groups

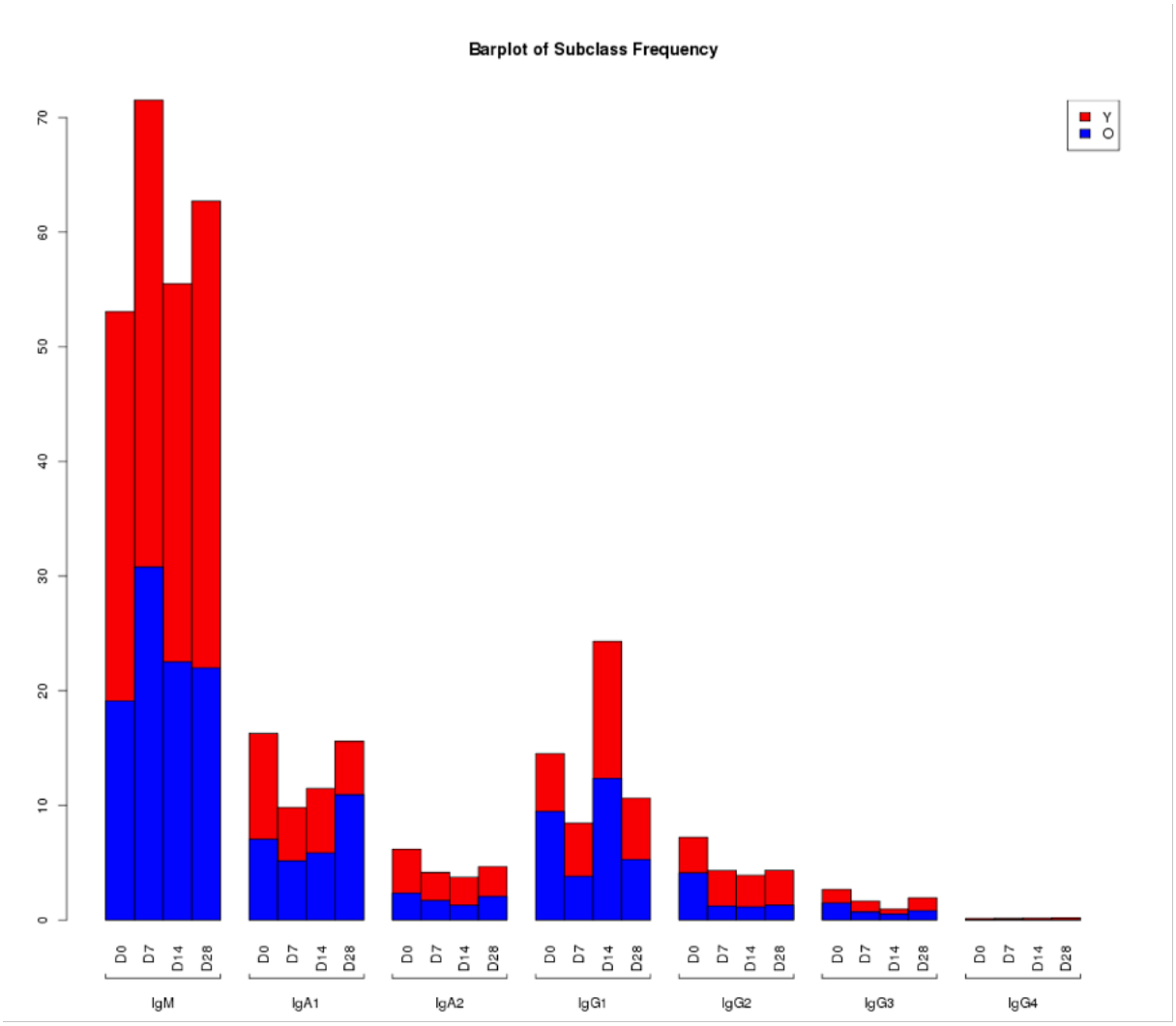

Figure 3. Bar plot of the subclass frequency between the age groups for different time points, including day 0, day 7, day 14, and day 28 post vaccination. The y-axis expresses percentage (%) of subclass frequency. The data from the old age group was collected from donors with the age 59, 56 and 55 and is presented as blue bars. The young age group included donors with the age of 34, 24 and 26 and is presented with red bars. The factors age and day was grouped and was normalized by all data in the sample. No transform data was used to generate the plot. The property to plot was set to “subclass”.

By comparing the frequencies on the y-axis, differences between the age groups can be observed in Figure 3. The comparison of frequency of subclasses on the y-axis concludes allows an estimation of the different frequencies. IgM was during day 0 approximately 55% for the young age group and 18% for the old age group, concluding the levels of IgM to be suppressed before the YF17D vaccine was administrated to the individuals. Day 7 post vaccination, the values changed to 40% for the young age group and 30% for the old age group. The young age group was shown to be 30% and the old 22% at day 14% post vaccination. The values for day 28 post vaccination were 40% for the young, and 20% for the old. The younger age group had higher frequencies of IgM than the old age group at all time points.

As visualized in Figure 3, the bar plot expresses the differences in immunoglobulin frequencies for different age groups and time points. The frequency of IgM was suggested to be increased 7 days post-vaccination for both age groups. The results indicate a predominant response of IgM, peaking at day 7 for both age groups. Following the induction of IgM, there is an indication of IgG1 and IgA1 secretion. The secretion of IgG1 peaked on day 14 post-vaccination for both age groups. The bar plot also suggests a low secretion of IgA2 and IgG2 and IgG3 but close to the non-existent level of IgG4.

Levels of IgG1 were potentially higher in the older age group at day 0 with approximately 10% in comparison to the younger age group with 5%. This is not caused by the vaccine but is a naturally occurring difference between the groups and is due to endogenous factors. The differences are visually comparable, but no statistical test can provide evidence against the null hypothesis and confirm the differences.

The Kidera Factors are used as measurements of the physicochemical properties of the sequences and can help us predict how certain compounds will act or behave functionally (Epstein et al., 2014). Distribution analysis performed in BRepertoire pointed out differences in the data distributions between the age groups for Kidera Factor 1, Kidera Factor 4, Kidera Factor 5, Kidera Factor 8, and Kidera Factor 9. The most significant differences were shown to be within the CDR3 properties of Kidera Factor 4 and Kidera Factor 9, distributions for the analyzed datasets. The P-values provided by the T-test on the distribution analysis of Kidera Factor 4 (expressed in Table 3) and Kidera Factor 9 (expressed in Table 4) implies differences in the CDR3 properties between the age groups that can imply a decline in immune response with age. These physicochemical properties play an important role in regulating the flexibility of the variable region and therefore may regulate their effective binding to the antigen (Epstein et al., 2014). Kidera Factor 4 is a measurement of the hydrophobic properties of the sequence and Kidera Factor 9 is a measurement of the pK.C Value of the peptide at the C-terminus of the fragment. Potential differences identified at time point 0 indicate a difference in B-cell repertoire or physicochemical properties due to endogenous factors and exclude the effect of the vaccine. These differences may be of interest for future research to identify pre-existing differences pre-vaccination.

The P-values provided by the T-test on the distribution analysis imply the difference in the CDR3 properties between the age groups. T-tests were performed in order to calculate the significance of the difference in data. The P-values generated from the T-Tests expressed in Table 3 and Table 4 indicated a significant difference in Kidera Factor 4 and 9 for day 7 and day 14 post-vaccination. Figure 4 illustrates the differences in the physicochemical properties of Kidera Factor 4 and Kidera Factor 9 as a distribution analysis of day 14 post-vaccination.

The differences in data distribution of Kidera Factor 4 expressed in Figure 4 suggests a skewing of the old age group data to the right, indicating a more hydrophobic CDR3 amino acid sequence in the older age group in comparison to the younger age group. This is confirmed with P-values for day 7, day 14, and day 28 post-vaccination (Table 3). The distribution of the physicochemical property Kidera Factor

9 data collected from the old age group is shown to be skewed to the left in Figure 4. The results indicate a difference in pK value at the C-terminus of the peptide fragment, suggesting an alteration in amino acids in the peptide fragments between old and young individuals.

Distribution Analysis of Kidera Factor 4 and Kidera Factor 9 day 14 post-vaccination.

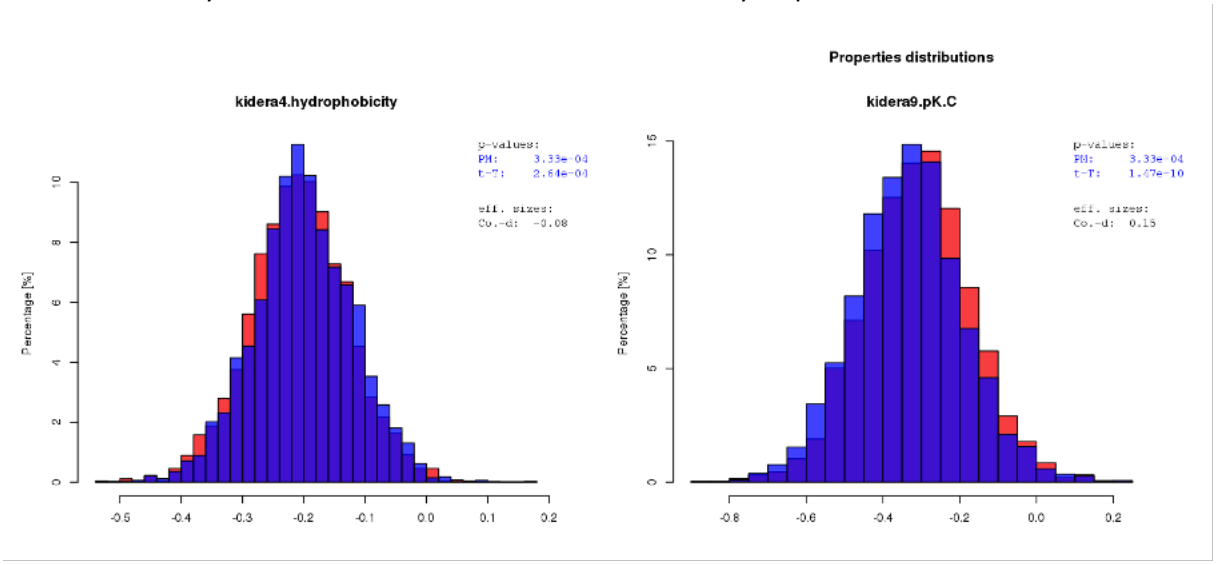

Figure 4. Distribution analysis for the physicochemical properties Kidera Factor 4 to the left, and hydrophobicity and Kidera Factor 9, pC.K value to the right. The distribution analysis expresses the difference in data between the age groups at day 14 post-vaccination. The dark blue area in the distribution analysis expresses the overlapping data between the age groups. The lighter blue bars presented data collected from the old age group donors of the age 59, 56, and 55, and the red bars represented data collected from the young age group of donors with the age of 34, 24, and 26. A T-test test was performed as a statistical test for Kidera Factor 4 and Kidera Factor 9, expressed in Table 3 and Table 4 respectively. The threshold for significance was set to P-value > 0.05 and the number of iterations was set to 3,000. The effect size was set to cohen's d and the confidence interval was set to 90%.

Table 3. P-Values for Kidera Factor 4 properties of the CDR3 on the variable region of the heavy chain.

Day	T-Test
0	0.175957382053371
7	0.0217857536612018
14	0.00026361546326006
28	0.000334267226126619

The P-values for the distribution analysis of the Kidera Factor 4 data expressed in Table 3 provide evidence for significant differences in the physicochemical property between the age groups. There was a significant difference on day 7, day 14, and day 28 in hydrophobicity between the age groups. The skewing of data indicates more extensive hydrophobic properties in the CDR3 of the old age group than at the young age group.

Table 4. P-Values for Kidera Factor 9 properties of the CDR3 on the variable region of the heavy chain.

Day	T-Test
0	0.232429759151108
7	0.045786342835387
14	1.47242210951222e-10
28	0.219983524724372

The P-values expressed in Table 4 provide proof of measurable differences in pK value at the C-terminus of the peptide fragment between the age groups on day 7 and day 14. The data presenting

the old age group was skewed to the left, indicating a more negatively charged composition of amino acids in the peptide fragment. Previous studies suggest cysteines to be able to change the physicochemical properties of CDR3 due to its ability to form disulfide bonds (Dunn-Walters et al., 2018). Table 3 and Table 4 conclude a significant difference in Kidera Factor 4 for day 7, day 14, and day 28 and in Kidera Factor 9 at day 7 and day 14 post-vaccination.

The gene frequency was examined aiming to identify differences in the use of genes between the age groups. BRepertoire generated a gene usage frequency box that presented the usage of the different monitoring of the V, D, and J segments (Epstein et al., 2014). The parameters "Vfamily", "Dfamily" and "Jfamily" were selected and compared in order to identify differences in frequency of usage. The generated plots did not indicate a difference in gene frequency among the age groups. There were no observed differences in the gene usage frequency in the IGHV gene, which agrees with previous studies (Davydov et al., 2018).

The length of the CDR3 amino acid sequence has been shown to play a central role in regulating the flexibility of the variable region of the antibody (Dondelinger et al., 2018). The dataset consisted of sequences of the heavy chain's CDR3 and was investigated in BRepertoire in order to examine the structure and identify potential differences between the age groups.
Length of CDR3 amino acid sequence

Box Whisker Plot of the CDR3 amino acid sequence length

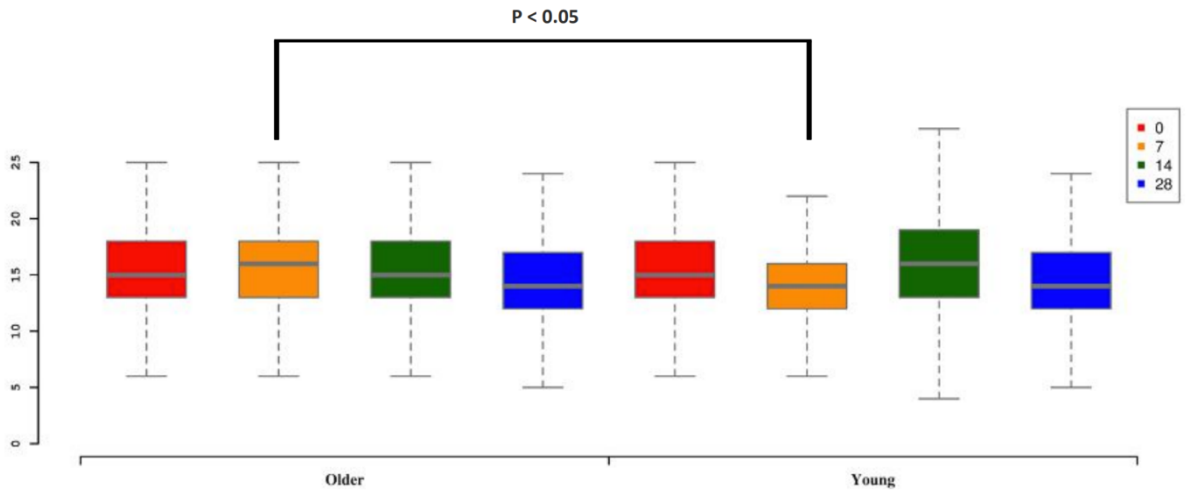

Figure 5. Box-Whisker plot generated in BRepertoire, represents numerical data of the CDR3 amino acid sequence length between the age groups at different time points, including day 0, day 7, day 14, and day 28 post-vaccination. The y-axis presents the number of amino acids, and the x-axis presents the different age groups. The length of the CDR3 amino acid sequence during the different time points is expressed as red (day 0), yellow (day 7), green (day 14) and blue (day 28). A two-tailed Mann-Whitney-Wilcoxon test was added as a statistical method and the threshold for significance is 0.05.

The yellow boxes shown in Figure 5 are representing the CDR3 amino acid sequence length at day 7 and are expressing differences between the age groups. A t-test was added to the Box-Whisker plot analysis in order to generate evidence about significance. The P-values generated indicated a significant difference in length of the CDR3 amino acid sequence length at day 7 post-vaccination between the age groups with a P-value of 0.00462. The amino acid sequence length of the CDR3 was not significantly different at the remaining time points including day 0, day 14, and day 28. The data

expressed in Figure 5 concludes the CDR3 amino acid sequence length to be significantly longer for the older age group at day 7 post-vaccination.

A model of antibody sequences originating from different donors was generated in PyMol in order to enable the structures to be visualized.

Model of aligned antibody structures.

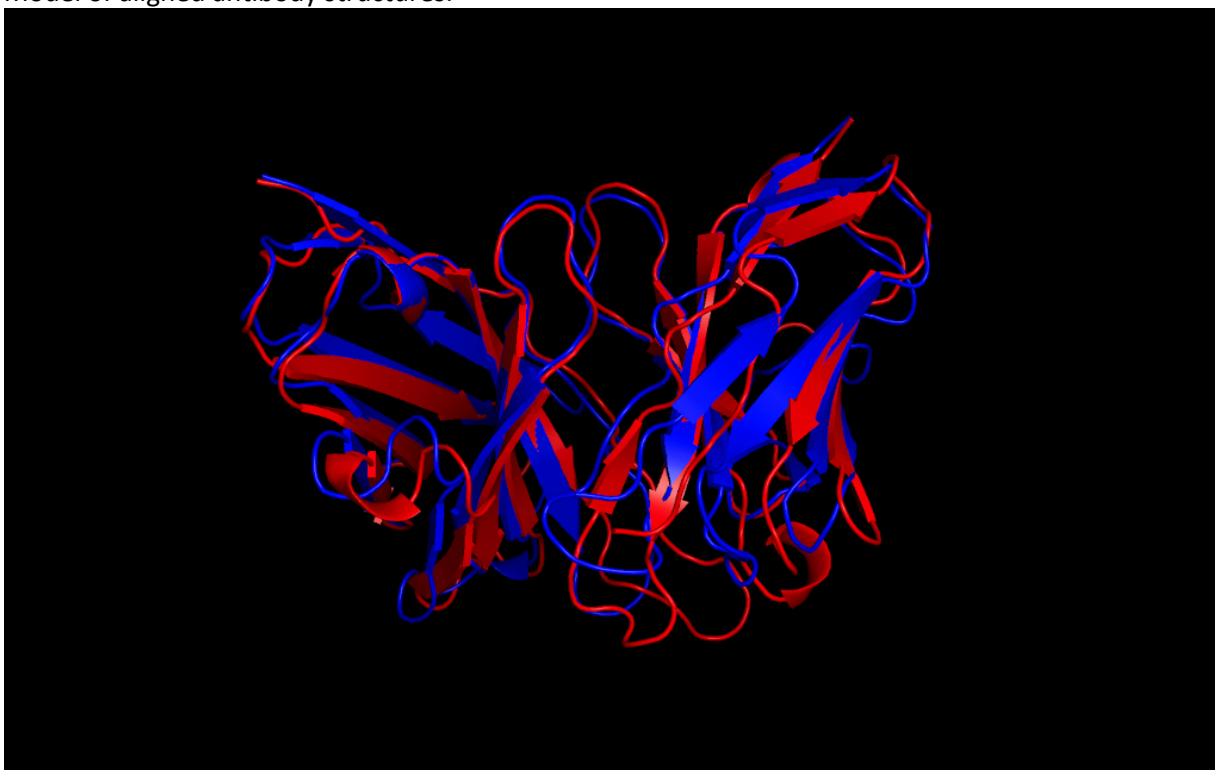

Figure 6. Antibody sequences of paired heavy and light chains aligned, originating from different donors of different ages. The red structure consists of a paired sequence originating from a 24-year-old donor and the blue structure represents a sequence originating from a 59-year-old donor. The chains were paired in ABodyBuilder and combined in PyMol to enable the visualization of the antibody structures. The confidence threshold in AbodyBuilder was set to 75%.

The aligned antibody sequences visualized in Figure 6 do not visually express differences in structure but provides an overview of the structure similarities. The results do not necessarily mean there is no difference in sequences.

Discussion

The adaptive immune response alters with age and will impact its ability to generate immunity to an antigen (Dunn-Walters & Ademokun, 2010). This report accounts for a variety of different factors contributing to the decline in immune response with age, including both physicochemical properties and subclass frequency.

Previous studies pointed out a quick humoral response to YF17 Vaccination. The predominant response includes an instant secretion of IgM, peaking at day 7 post-vaccination and will persist for approximately three months. Following the secretion of IgM, there is a slower but more consistent secretion of IgG. The IgM will decline eventually while the IgG provides long-term protection against the antigen since it lasts for years (Domingo et al., 2018). These conclusions are in line with the results presented in Figure 3. Figure 3 accounts for a high level of IgM secretion, peaking at day 7 post-vaccination. The initial response to the vaccine triggers the secretion of IgM peaking between day 4 and 7 post-vaccination (Plotkin et al., 2018), which agrees with the generated results of the subclass frequency presented in Figure 3. During day 4 to 7 post-vaccination the IgG levels are expected to be >16 - fold lower than IgM (Plotkin et al., 2018), which is backed up by the results.

Merbah et al. studied the humoral response including the subclass frequency as a response to YF17 vaccination in 2018. They pointed out the different immunoglobulin levels in order to further identify the full humoral response to the vaccine. The majority of the secreted IgG was shown to be IgG1, followed by IgG2 and IgG3 respectively (Merbah et al., 2020). This is confirmed by the levels of IgG1, IgG2 and IgG3 illustrated in Figure 3. The subclass IgG4 is not identified, tentatively due to the absence of IgG4 in the humoral immune response. Alterations in the frequency of the subclasses IgG1 and IgA1 were observed. The usage of IgA2 and IgG2 was shown to increase with age (Ghraichy et al., 2020). The results expressed in Figure 3 suggest a slight difference in the levels of immunoglobulin between the age group. In general, the young age group presented as red bar plots have higher levels of immunoglobulins, with the exemption for IgG1 at day 0 for the older age group. IgG1 is the predominant isotype and accounts for approximately 70% of the circulating IgG levels and is responsible for mediating a variety of different complement-mediated functions within the immune system. A lack of IgG1 may cause abnormalities and dysfunctions in pathways regulating the immune response (Napodano et al., 2020).

Some observed differences between the age groups for day 0 presented in Figure 3 are caused by endogenous or exogenous factors separated from the YF17 vaccination. This includes an increase in IgG1 at day 0 post-vaccination and decreased IgM levels at day 0 in comparison to the young age group. These factors are of interest for the basic research in how the adaptive immune response alters with age and might be a target for future drug discovery and development or improvement of the vaccine regimen. The adaptive immune response in elderly patients has been studied to identify a decline in IgM, which agrees with the results generated in BRepertoire presented in Figure 3. The quantity of immune cells stays constant with age, but the quality of the immune response is suppressed since the antibodies produced are less specific to the antigens (Dunn-Walters et al., 2020).

Results generated in previous studies indicate a more diverse immune response in the younger patient group, which is suggested to be attributed to a higher efficient somatic hypermutation mechanism. Elderly patients were shown to have a decreased response to both primary and secondary shots of attenuated virus vaccines. The decreased immune response might cause a lack of protection from the virus (Davydov et al., 2018).

Other studies made out of observations of the immune response to other YF17D vaccines identified a decrease in neutralizing properties of the antibodies at day 14 in older patients, which is suggested to

be partly because of the decreased CD8Tcell response. This points towards a compromised effector phase for older patients. The CD4 T cells were shown to have similar antigen affinity between age groups, but they were shown to have a decreased quantitative polyfunctionality in comparison with younger age groups. The affinity for the yellow fever virus antigen of CD4 T cells was significantly decreased with time, pointing towards a decline in protection against the antigen (Gustafson et al., 2020).

The physicochemical properties of antibodies can be investigated by examining the Kidera Factors encoded by the studied sequences. The Kidera factors have been derived to encode 188 physical properties of the 20 amino acids using dimension reduction techniques. They consist of a 10-dimensional vector of orthogonal factors for each amino acid (Nakai et al., 1988). Differences in physicochemical properties point out differences contributing to the alteration with age, which can be explained by the impact of the amino acids and the length of the sequence. These factors are able to detect structural properties of the variable region, and hence, affect the antigen affinity of the variable region (Schroeder & Cavacini, 2010). Structural changes in the variable region affecting its affinity for antigens have been observed in the same variable region on IgA2 compared to IgG1, suggesting different subclasses to have a variety of repertoires of the variable region (Su et al., 2018).

The Kidera Factors of the B-cell repertoire originating from the different age groups were investigated in order to classify the physicochemical properties of the CDR3, to expand the understanding of how their properties will affect the binding properties of the variable region (Laffy et al., 2017). Distribution analysis performed in BRepertoire pointed out skewing in data between the age groups. The most significant differences were found in Kidera Factor 4 and Kidera Factor 9, concluding these to be of high interest in order to target differences in the B-cell repertoire between age groups.

Kidera Factor 4 measures the hydrophobic properties of the antibody. Antibody specificity and affinity are highly dependent on the hydrophobic properties, concluding alterations to impact the structure and function of the antibody (Davydov et al., 2018). Antibodies with more hydrophobic properties are at higher risk of interacting with amino acids centrally located on the CDR3 which may cause the antibody to cross-react with unwanted residues. The more hydrophobic properties, the higher risk of cross-reacting and unspecific binding (Davydov et al., 2018). The left distribution analysis in Figure 4 illustrates the data distribution of Kidera Factor 4 at day 14 post-vaccination. The data collected from the B-cell repertoire of the old age group is skewed to the right, indicating more hydrophobic properties at the old age group than in the young age group. The P-values expressed in Table 3 confirm the differences in hydrophobic properties of the CDR3 to be significantly different between the age group. These results point out a potential cause of the decline in immune response with age due to the increased risk of cross-reacting and nonspecific antigen binding (Davydov et al., 2018).

Kidera Factor 9 is a measurement of the pK value of the amino acids present in the sequence of the C-terminus of the peptide fragment (Williams et al., 2013). Measuring the charge of the amino acids contributes to an estimation of the stability and functionality of the protein. The charge of the amino acid regulates the interaction with other molecules and affects the stability and structure of the antibody. Missense mutations are mutations that induce an exchange in amino acids produced by the ribosome due to a sequence alteration. These mutations may cause errors in protein production if the mutation induces changes that alter the amino acid composition. Depending on the extent of the mutation, this may impact the structure of the protein (Yampolsky & Stoltzfus, 2005). Mutations that induce positively charged amino acids will cause a decrease in the pK-value of the C-terminus, and a mutation that introduces a negatively charged amino acid in the sequence of the peptide will result in an increased pK value (Grimsley et al., 2008). Disruptions in pK value can cause alteration in functionality, which could contribute to the decline in the immune response. Significant changes might even cause the antibodies to aggregate or misfold (Grimsley et al., 2008).

There is a limited amount of available research observing differences in Kidera Factor 9 between B-cell repertoires originating from different age groups. The distribution analysis to the right in Figure 4 visualizes a skewing of the old age group B-cell repertoire data to the left, indicating a more negative pK value of the amino acids in the peptide of the C-terminus at the old age group donors. The P-values expressed in Table 4 confirms a significant difference in pK value at day 7 and 14 post-vaccination, which points towards an endogenous alteration caused by the YF17D vaccine. The difference in the pK value of the CDR3 amino acid sequence between the age group is observed at the time point where the immune response is expected to peak (Plotkin et al., 2018). The skewing of Kidera Factor 9 data presented in Figure 4 contributes to the hypothesis that there is an alteration in amino acid composition at the C-terminus peptide sequence. Identification of the alterations in the sequence is of high interest in order to evaluate the impact of certain missense mutations. Current research points towards an increased risk of unspecific interactions due to a mutation that induces arginine which is a positively charged amino acid. Antibody sequences lacking arginine were shown to have a decreased risk of unspecific binding interactions. The polar amino acid serine on the other hand was shown to be unable to induce unspecific binding due to its hydrophobicity. The same study also concluded tyrosine and serine combined to exclusively be the only mutations occurring in the complementary determining regions to decrease the risk of unspecific interactions. This has been identified as a target of interest for future research of the sequence and structure of the antibody (Rabia, 2018).

The differences between the B-cell repertoire of the age groups in physicochemical properties observed in Kidera Factor 4 and Kidera Factor 9 contributes to a fundamental identification of how the physicochemical differences of CDR3 alter with age. This could be of interest for a potential drug or antibody therapy for future perspectives and provides a fundamental research basis of how the alterations in properties and structure affect antigen affinity, interaction ability, and structure of the CDR3 (Laffy et al., 2017).

The length of the CDR3 amino acid sequence was investigated in order to reveal differences in the effect of the vaccine on the B-cell repertoire between the age groups. The length of the CDR3 sequence varies between 5-15 amino acids (Martin et al., 2003). The results presented in Figure 5 agrees with the previous studies of the CDR3 amino acid sequence length, where CDR3 amino acid sequences included in the investigations had lengths of 5-18 amino acids. Alterations in the CDR3 structure impact the rearrangement of the genes, which adapts the variable regions' affinity for antigens (Chlewicki et al., 2005). The physicochemical properties of the CDR3 region have shown significant differences which impacts the interaction abilities. Furthermore, this affects the probability for cross-reactions and unspecific binding (Davydov et al., 2018). Identifying differences in the CDR3 is of high value due to its extensive impact on the antigen-binding of the variable region. The CDR3 regions are the most variable part of the antibody since differences in the coding regions cause an extensive diversity in structures (Dunn-Walters et al., 2018). The length of the CDR3 amino acid sequence will impact the flexibility of the variable region and hence, affect the antigen affinity (Ovchinnikov et al., 2018).

The results of the BRepertoire analysis of the numerical data of the CDR3 amino acid sequence length is presented in Figure 5 and is indicating an increase in CDR3 amino acid sequence length at day 7 post-vaccination for the old age group. The results are confirmed by the p-value = 0.00462, indicating a statistical difference in CDR3 amino acid sequence length at day 7 post-vaccination between the old and young age group. The results presented in Figure 5 are in line with previous studies of the CDR3, concluding the amino acid sequence to be increased in older individuals. Unlike the results presented in Figure 5, the baseline CDR3 has been shown to be increased in length in general for the older patient group at all time points (Ademokun et al., 2011). The alteration in CDR3 amino acid sequence length in the older patient group is one of the suggested reasons for the decline in immune response, due to

the effect on flexibility and structure which impacts the antigen affinity and binding properties of the variable region (Barrios et al., 2004).

Previous research also observed a positive correlation between Kidera Factor 4 (hydrophobicity) and CDR3 amino acid sequence length. The older age group had both increased hydrophobic properties and longer CDR3 fragments, and the younger age group less hydrophobic and shorter CDR3 amino acid sequences (Ademokun et al., 2011). To conclude, the physicochemical properties of the CDR3 in the B-cell repertoires from the older age group had longer and more hydrophobic peptide fragments, which is in line with the results presented in Figure 4 and Figure 5 (Davydov et al., 2018). The alteration in properties causes the flexibility and structure of the CDR3 to change, causing disruptions in the antigen affinity which contributes to the decline in immune response with age (Ghraichy et al., 2020).

Class switching recombination from IgM to IgG and IgA relies on the B-cells ability to differentiate. The cells ensure a functional arrangement of genes by observing if the combination of segments resulted in the production of a specific protein, AID (Activation-induced cytidine deaminase). This protein is responsible for the transcription of DNA which enables the RNA to be translated into a protein. Failure to produce the functional arrangement of genes will result in apoptosis (Chi et al., 2020). AID has been confirmed *in vitro* to decline with age, which may limit the DBS (Frasca et al., 2020). The amount of DBS has been identified as one of the factors affecting CSR since it's crucial for the recombination of the S regions of the B-cell sequence (Stavnezer et al., 2008). Class switch recombination is dependent on double-strand breaks (DBS) present in the regions in order to introduce replacement of exons.
The enzyme AID can be used as a biomarker to identify the immune response to an antigen since it's correlated with the antibodies produced in response to the infection (Frasca et al., 2020). The expression of the enzyme AID has been confirmed to be suppressed in older individuals. Since AID regulates the class switching recombination, this limits the production of IgG in older patients. Lack of IgG secretion itself contributes to a decline in immune response since it makes the host cells more susceptive to antigens (Napodano et al., 2020). The suppressed expression of AID could act as a potential drug target for a future drug discovery and development process.

The identification of the differences in the immune response to the YF17 vaccine contributes to an improved understanding of the altered immune response with age which enables more extensive research of the vaccine regimen in order to ensure the efficacy and safety of vaccines. An altered immune response with age will increase the frequency of adverse events to the vaccine in the older age group, in comparison to the younger patients. This is suggested because of the decreased ability to have an efficient immune response to the antigen (Plotkin et al., 2018).

The currently available vaccines are based on the T follicular helper cells and the B cells to secrete specified antibodies in order to generate immune response. By integrating the existing research of both B-cell and T-cell response to the vaccine, collects highly valued information required for enabling a vaccine regimen improvement and understanding the alteration in the adaptive immune system with age (Gustafson et al., 2020). The primary response to the YF vaccine was studied by Gustafson et al., and concluded older patients to have a suppressed secretion of antibodies in comparison to younger patients, most significant at day 5 post-vaccination, 28 days post-vaccination, the levels of antibodies present was equal. Gustafson et al., concluded the immune system in older patients to still have the capacity of generating immunity to the vaccine despite the delay in immune response, concluding the immunity to be postponed and supressed.

BRepertoire enables the physicochemical properties of the CDR3 to be revealed, which contributes to an improved therapeutic monoclonal antibody research discovery and development (Laffy et al., 2017). Currently, monoclonal antibodies account for over 50% of proteins used as therapeutics on the

market (Dennler et al., 2015). Identifying the properties of the antibodies is crucial to predict how they will behave and interact with the target cells (Chiu et al., 2019).

The lack of sufficient age diversity in the dataset limits examination of the physicochemical properties of the B-cell repertoires. A more extensive donor collection with a wider range of age would possibly result in more extensive and distinguishable results and generate a more reliable outcome.

Ethical aspects

The samples examined during this project was collected from healthy volunteers. The identities of the donors were anonymized, which excluded the risk of compromising the integrity for the donors. In addition, the samples collected was blood samples which reduced the risk of the donors having an unpleasant experience.

Bioinformatic tools can be used in drug discovery development, toxicology predictions, and basic research applications by integrating the *in silico* results from analyzing transcriptomes. This replaces and reduces the use of animals by presenting methods to test safety and efficacy *in vitro* instead of using animals. Using bioinformatic software to identify biomarkers are expanding in clinical applications currently, and recent studies used a bioinformatic approach for identifying an upregulation in a gene indicating breast cancer. Bioinformatics was also used in order to compare, analyze and evaluate the results from different therapeutic methods in order to cure or treat the cancer (Yang et al., 2019).

The management of extensive sets of data to use as reference genomes are of high value for diagnostic procedures. Bioinformatic tools are effective for examining genomic data in order to identify difference in immune response that could be related to underlying pathologies. Highlighting these differences may increase the chances of providing the patient with the right treatment which increases the survival rate and therefore improves the quality of life for the patients. Since the accumulation of data ensures an efficient and reliable diagnosis, this also decreases the risk of miss-diagnosing or miss-treating patients (Ienca et al., 2018). At a global health-economic perspective, discovering the risk of developing disease early cuts the costs of treatments. The bioinformatic application to healthcare is extensive since it allows accumulation of genomic data (Ienca et al., 2018). This provides crucial references for the development of new diagnostics such as biomarkers, therapeutics and vaccines (Singh, 2021).

Bioinformatics does not only hold great promise in facilitating diagnostics procedures but also contributes to a more cost and time efficient drug discovery and development process (Singh, 2021). The process of drug discovery and development is expensive and highly time consuming. Bioinformatics can be used to predict the effect of the therapeutic on a cell and therefor cut the price for the pre-clinical trials. An improved software application to *in vitro* work will also contribute to the 3R principle involving reducing, refine and replace animal use in research. The results might even provide a better analysis of how the mechanism works on human cells, since all results generated in animal models may not be applicable to the human mechanism of action (Grafström et al., 2015). The examined B-cell repertoire provides an improved understanding of the transcriptome of B-cells, and its role in CSR which may be applicable to other types of research involving the immune response (Attaf et al., 2020).

Bioinformatics enabled a revolutionizing improvement of techniques used to discover inherited diseases patterns. Ethics should always be considered when analyzing confidential private patient data, since revealing the genomic data also opens up the dilemma to distinguish between what information is harmful or helpful (Ienca et al., 2018). Unexpected discoveries of pathological characteristics open up a dilemma of evaluating the genetic information. In some cases, a pessimistic

result is unjustified to communicate to the patient in cases where the disease progression is uncertain or at low risk, while in some cases might be lifesaving to inform the patient in time and ensure accurate treatment of the disease in time. Bioinformatics can and will also be used to evaluate a patient's risk of developing disease by analyzing genomic data. Discovering a disease early in the progression increases the survival rate and has an extensive impact on society by reducing stress associated with a diagnosis for both the patient and its relatives (Wang, 2017). The accumulation of genomic data comes with the risk of compromising the data privacy if unauthorized individuals have the ability to access it. Thus, it is important to carefully store and manage the patient data in order to ensure safety for the patient's confidential data (Ienca et al., 2018).

Conclusion

The identified differences in physicochemical properties of the CDR3s studied include length, hydrophobicity and pK-value of the peptide fragment at the C-terminus. Differences in subclass frequency was observed and concluded the older age group to have a suppressed secretion of IgM, which decreases the efficacy of the predominant immune response to an antigen. A decline of AID levels in older patients could also play a role in the suppressed immune response. These factors are suggested to contribute to the decline in immune response with age due to their ability to alter the structure and antigen affinity of the variable region of the antibody. By identifying what factors contribute to a decline in immune response, one could target potential factors playing a role in an insufficient immune response system, caused by different endogenous and/or exogenous factors.

Acknowledgements

Project funded by King's College London and University of Skövde.

I am grateful for the opportunity to participate in the project at King's College London and for all the support I received throughout the development of my thesis dissertation. First of all, I would like to thank my supervisor Prof. Franca Fraternali for supporting, guiding, and leading me through the project. You introduced the world of bioinformatics to me and opened up a new and broader horizontal perspective on science. Second of all, I would like to thank Joseph Chi-Fung Ng for all the guidance through the learning of the bioinformatics software tools and the valuable input on my presentations. I would also like to thank Dongjun Guo for all the help and support. In addition, I would like to thank Deborah Dunn-Walters, Luke Marques, Alexander Stewart, Emma Sinclair, and Timir Weston for all the help. Last but not least, many thanks to the whole team for everything you taught me in your presentations and stimulating discussions.

Thanks to the University of Skövde and King's College London for giving me the chance to participate in the project. I am beyond grateful for this experience and the opportunity to get an insight into the bioinformatic research. Many thanks to Sanja Jurcevic for encouraging and inspiring me, and thanks to the international coordinator Frida Lindgren for supporting and helping me through the exchange.

References

Ademokun, A., Wu, Y., Martin, V., Mitra, R., Sack, U., & Baxendale, H. et al. (2011). Vaccination-induced changes in human B-cell repertoire and pneumococcal IgM and IgA antibody at different ages. *Aging Cell*, *10*(6), 922-930. https://doi.org/10.1111/j.1474-9726.2011.00732.x

Alamyar, E., Duroux, P., Lefranc, M., & Giudicelli, V. (2012). IMGT® Tools for the Nucleotide Analysis of Immunoglobulin (IG) and T Cell Receptor (TR) V-(D)-J Repertoires, Polymorphisms, and IG Mutations: IMGT/V-QUEST and IMGT/HighV-QUEST for NGS. *Methods In Molecular Biology™*, 569-604. doi: 10.1007/978-1-61779-842-9_32

Attaf, N., Cervera-Marzal, I., Dong, C., Gil, L., Renand, A., Spinelli, L., & Milpied, P. (2020). FB5P-seq: FACS-Based 5-Prime End Single-Cell RNA-seq for Integrative Analysis of Transcriptome and Antigen Receptor Repertoire in B and T Cells. *Frontiers In Immunology*, *11*. doi: 10.3389/fimmu.2020.00216

Barrios, Y., Jirholt, P., & Ohlin, M. (2004). Length of the antibody heavy chain complementarity determining region 3 as a specificity-determining factor. *Journal Of Molecular Recognition*, *17*(4), 332-338. https://doi.org/10.1002/jmr.679

Bransteitter, R., Pham, P., Scharff, M., & Goodman, M. (2003). Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. *Proceedings Of The National Academy Of Sciences*, *100*(7), 4102-4107. https://doi.org/10.1073/pnas.0730835100

Chi, X., Li, Y., & Qiu, X. (2020). V(D)J recombination, somatic hypermutation and class switch recombination of immunoglobulins: mechanism and regulation. *Immunology*, *160*(3), 233-247. doi: 10.1111/imm.13176

Chiu, M., Goulet, D., Teplyakov, A., & Gilliland, G. (2019). Antibody Structure and Function: The Basis for Engineering Therapeutics. *Antibodies*, *8*(4), 55. doi: 10.3390/antib8040055

Chlewicki, L., Holler, P., Monti, B., Clutter, M., & Kranz, D. (2005). High-affinity, Peptide-specific T Cell Receptors can be Generated by Mutations in CDR1, CDR2 or CDR3. *Journal Of Molecular Biology*, *346*(1), 223-239. doi: 10.1016/j.jmb.2004.11.057

Chung, E. (2014). Vaccine allergies. *Clinical And Experimental Vaccine Research*, *3*(1), 50. https://doi.org/10.7774/cevr.2014.3.1.50

Davis, E., Beck, A., Strother, A., Thompson, J., Widen, S., & Higgs, S. et al. (2019). Attenuation of Live-Attenuated Yellow Fever 17D Vaccine Virus Is Localized to a High-Fidelity Replication Complex. *Mbio*, *10*(5). doi: 10.1128/mbio.02294-19

Davydov, A., Obraztsova, A., Lebedin, M., Turchaninova, M., Staroverov, D., & Merzlyak, E. et al. (2018). Comparative Analysis of B-Cell Receptor Repertoires Induced by Live Yellow Fever Vaccine in Young and Middle-Age Donors. *Frontiers In Immunology*, *9*. doi: 10.3389/fimmu.2018.02309

Dennler, P., Fischer, E., & Schibli, R. (2015). Antibody Conjugates: From Heterogeneous Populations to Defined Reagents. *Antibodies*, *4*(3), 197-224. doi: 10.3390/antib4030197

Domingo, C., Charrel, R., Schmidt-Chanasit, J., Zeller, H., & Reusken, C. (2018). Yellow fever in the diagnostics laboratory. *Emerging Microbes & Infections*, *7*(1), 1-15. https://doi.org/10.1038/s41426-018-0128-8

Dondelinger, M., Filée, P., Sauvage, E., Quinting, B., Muyldermans, S., Galleni, M., & Vandevenne, M. (2018). Understanding the Significance and Implications of Antibody Numbering and Antigen-Binding Surface/Residue Definition. *Frontiers In Immunology*, *9*. https://doi.org/10.3389/fimmu.2018.02278

Dunbar, J., Krawczyk, K., Leem, J., Marks, C., Nowak, J., & Regep, C. et al. (2016). SAbPred: a structure-based antibody prediction server. *Nucleic Acids Research*, *44*(W1), W474-W478. doi: 10.1093/nar/gkw361

Dunn-Walters, D., & Ademokun, A. (2010). B cell repertoire and ageing. *Current Opinion In Immunology*, *22*(4), 514-520. doi: 10.1016/j.coi.2010.04.009

Epstein, M., Barenco, M., Klein, N., Hubank, M., & Callard, R. (2014). Revealing Individual Signatures of Human T Cell CDR3 Sequence Repertoires with Kidera Factors. *Plos ONE*, *9*(1), e86986. doi: 10.1371/journal.pone.0086986

Ewer, K., Barrett, J., Belij-Rammerstorfer, S., Sharpe, H., Makinson, R., & Morter, R. et al. (2020). T cell and antibody responses induced by a single dose of ChAdOx1 nCoV-19 (AZD1222) vaccine in a phase 1/2 clinical trial. *Nature Medicine*. https://doi.org/10.1038/s41591-020-01194-5

Figueiredo, C., Matsuo, A., Massaoka, M., Polonelli, L., & Travassos, L. (2014). Anti-tumor activities of peptides corresponding to conserved complementary determining regions from different immunoglobulins. *Peptides*, *59*, 14-19. doi: 10.1016/j.peptides.2014.06.007

Frasca, D., Blomberg, B., Garcia, D., Keilich, S., & Haynes, L. (2020). Age-related factors that affect B cell responses to vaccination in mice and humans. *Immunological Reviews*, *296*(1), 142-154. https://doi.org/10.1111/imr.12864

Frasca, D., Diaz, A., Romero, M., & Blomberg, B. (2016). The generation of memory B cells is maintained, but the antibody response is not, in the elderly after repeated influenza immunizations. *Vaccine*, *34*(25), 2834-2840. https://doi.org/10.1016/j.vaccine.2016.04.023

Geisberger, R., Lamers, M., & Achatz, G. (2006). The riddle of the dual expression of IgM and IgD. *Immunology*, *0*(0), 060526021554006. https://doi.org/10.1111/j.1365-2567.2006.02386.x

Gemenetzi, K., Agathangelidis, A., Zaragoza-Infante, L., Sofou, E., Papaioannou, M., Chatzidimitriou, A., & Stamatopoulos, K. (2020). B Cell Receptor Immunogenetics in B Cell Lymphomas: Immunoglobulin Genes as Key to Ontogeny and Clinical Decision Making. *Frontiers In Oncology*, *10*. doi: 10.3389/fonc.2020.00067

Gershman, M., Staples, J., Bentsi-Enchill, A., Breugelmans, J., Brito, G., & Bastos Camacho, L. et al. (2012). Viscerotropic disease: Case definition and guidelines for collection, analysis, and presentation of immunization safety data. *Vaccine*, *30*(33), 5038-5058. https://doi.org/10.1016/j.vaccine.2012.04.067

Ghraichy, M., Galson, J., Kovaltsuk, A., von Niederhäusern, V., Schmid, J., & Recher, M. et al. (2020). Maturation of the Human B-Cell Receptor Repertoire with Age. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3518535

Gibney, K., Kosoy, O., Fischer, M., Edupuganti, S., Lanciotti, R., & Delorey, M. et al. (2012). Detection of Anti-Yellow Fever Virus Immunoglobulin M Antibodies at 3–4 Years Following Yellow Fever Vaccination. *The American Journal Of Tropical Medicine And Hygiene*, *87*(6), 1112-1115. https://doi.org/10.4269/ajtmh.2012.12-0182

Grafström, R., Nymark, P., Hongisto, V., Spjuth, O., Ceder, R., & Willighagen, E. et al. (2015). Toward the Replacement of Animal Experiments through the Bioinformatics-driven Analysis of 'Omics' Data from Human Cell Cultures. *Alternatives To Laboratory Animals*, *43*(5), 325-332. doi: 10.1177/026119291504300506

Grimsley, G., Scholtz, J., & Pace, C. (2008). A summary of the measured pKvalues of the ionizable groups in folded proteins. *Protein Science*, NA-NA. https://doi.org/10.1002/pro.19

Gustafson, C., Kim, C., Weyand, C., & Goronzy, J. (2020). Influence of immune aging on vaccine responses. *Journal Of Allergy And Clinical Immunology*, *145*(5), 1309-1321. https://doi.org/10.1016/j.jaci.2020.03.017

Ienca, M., Ferretti, A., Hurst, S., Puhan, M., Lovis, C., & Vayena, E. (2018). Considerations for ethics review of big data health research: A scoping review. *PLOS ONE*, *13*(10), e0204937. https://doi.org/10.1371/journal.pone.0204937

Iwasaki, A., & Medzhitov, R. (2015). Control of adaptive immunity by the innate immune system. *Nature Immunology*, *16*(4), 343-353. https://doi.org/10.1038/ni.3123

Kotnis, A., Du, L., Liu, C., Popov, S., & Pan-Hammarström, Q. (2008). Non-homologous end joining in class switch recombination: the beginning of the end. *Philosophical Transactions Of The Royal Society B: Biological Sciences*, *364*(1517), 653-665. doi: 10.1098/rstb.2008.0196

Laffy, J., Dodev, T., Macpherson, J., Townsend, C., Lu, H., Dunn-Walters, D., & Fraternali, F. (2017). Promiscuous antibodies characterised by their physico-chemical properties: From sequence to structure and back. *Progress In Biophysics And Molecular Biology*, *128*, 47-56. https://doi.org/10.1016/j.pbiomolbio.2016.09.002

Li, T., Tracka, M., Uddin, S., Casas-Finet, J., Jacobs, D., & Livesay, D. (2015). Rigidity Emerges during Antibody Evolution in Three Distinct Antibody Systems: Evidence from QSFR Analysis of Fab Fragments. *PLOS Computational Biology*, *11*(7), e1004327. doi: 10.1371/journal.pcbi.1004327

Lieber, M. (2010). The Mechanism of Double-Strand DNA Break Repair by the Nonhomologous DNA End-Joining Pathway. *Annual Review Of Biochemistry*, *79*(1), 181-211. doi: 10.1146/annurev.biochem.052308.093131

Margreitter, C., Lu, H., Townsend, C., Stewart, A., Dunn-Walters, D., & Fraternali, F. (2018). BRepertoire: a user-friendly web server for analysing antibody repertoire data. *Nucleic Acids Research*, *46*(W1), W264-W270. https://doi.org/10.1093/nar/gky276

Martin, D., Bradl, H., Collins, T., Roth, E., Jäck, H., & Wu, G. (2003). Selection of Ig μ Heavy Chains by Complementarity-Determining Region 3 Length and Amino Acid Composition. *The Journal Of Immunology*, *171*(9), 4663-4671. https://doi.org/10.4049/jimmunol.171.9.4663

Martin, V., Wu, Y., Townsend, C., Lu, G., O'Hare, J., & Mozeika, A. et al. (2016). Transitional B Cells in Early Human B Cell Development – Time to Revisit the Paradigm?. *Frontiers In Immunology*, *7*. https://doi.org/10.3389/fimmu.2016.00546

Merbah, M., Wollen-Roberts, S., Shubin, Z., Li, Y., Bai, H., & Dussupt, V. et al. (2020). A high-throughput multiplex assay to characterize flavivirus-specific immunoglobulins. *Journal Of Immunological Methods*, *487*, 112874. https://doi.org/10.1016/j.jim.2020.112874

Ms. Vipula, & Ms. Atula. (2018). *Human Anatomy and Physiology: For Undergraduate Students of Pharmacy, Nursing, Physiotherapy and Other Paramedical Sciences: Vol. First edition*. Laxmi Publications Pvt Ltd. (page279-281)

Nakai, K., Kidera, A., & Kanehisa, M. (1988). Cluster analysis of amino acid indices for prediction of protein structure and function. "*Protein Engineering, Design And Selection*", *2*(2), 93-100. https://doi.org/10.1093/protein/2.2.93

Napodano, C., Marino, M., Stefanile, A., Pocino, K., Scatena, R., & Gulli, F. et al. (2020). Immunological Role of IgG Subclasses. *Immunological Investigations*, *50*(4), 427-444. https://doi.org/10.1080/08820139.2020.1775643

Ovchinnikov, V., Louveau, J., Barton, J., Karplus, M., & Chakraborty, A. (2018). Role of framework mutations and antibody flexibility in the evolution of broadly neutralizing antibodies. *Elife*, *7*. https://doi.org/10.7554/elife.33038

Parham, P. (2015). The Immune System (4th ed., pp. 1-10). Garland Science.

Parkin, J., & Cohen, B. (2001). An overview of the immune system. The Lancet, 357(9270), 1777-1789. doi: 10.1016/s0140-6736(00)04904-7

Plotkin, S., Orenstein, W., & Offit, P. (2018). *Plotkin's vaccines* (7th ed., pp. 1181-1265).

Rabia, L., Desai, A., Jhajj, H., & Tessier, P. (2018). Understanding and overcoming trade-offs between antibody affinity, specificity, stability and solubility. *Biochemical Engineering Journal*, *137*, 365-374. doi: 10.1016/j.bej.2018.06.003

Reis-Filho, J. (2009). Next-generation sequencing. *Breast Cancer Research*, *11*(S3). https://doi.org/10.1186/bcr2431

Schroeder, H., & Cavacini, L. (2010). Structure and function of immunoglobulins. *Journal Of Allergy And Clinical Immunology*, *125*(2), S41-S52. https://doi.org/10.1016/j.jaci.2009.09.046

Shi, Z., Zhang, Q., Yan, H., Yang, Y., Wang, P., & Zhang, Y. et al. (2019). More than one antibody of individual B cells revealed by single-cell immune profiling. Cell Discovery, 5(1). https://doi.org/10.1038/s41421-019-0137-3

Singh, H. (2021). Bioinformatics: Benefits to Mankind. *International Journal Of Pharmtech Research*, *4*, 242-248. Retrieved 7 May 2021, from.

Stavnezer, J., Guikema, J., & Schrader, C. (2008). Mechanism and Regulation of Class Switch Recombination. *Annual Review Of Immunology*, *26*(1), 261-292. https://doi.org/10.1146/annurev.immunol.26.021607.090248

Stratton, K., Ford, A., Wright Clayton, E., & Rusch, E. (2011). *Adverse Effects of Vaccines*. Institute of Medicine National Academies Press Washington (DC).

Su, C., Lua, W., Ling, W., & Gan, S. (2018). Allosteric Effects between the Antibody Constant and Variable Regions: A Study of IgA Fc Mutations on Antigen Binding. *Antibodies*, *7*(2), 20. doi: 10.3390/antib7020020

Tanabe, K., & Kajihara, R. (2020). The role of protein phosphorylation in the regulation of class switch recombination. *BIOCELL*, *44*(4), 545-558. https://doi.org/10.32604/biocell.2020.012740

Townsend, C., Laffy, J., Wu, Y., Silva O'Hare, J., Martin, V., & Kipling, D. et al. (2016). Significant Differences in Physicochemical Properties of Human Immunoglobulin Kappa and Lambda CDR3 Regions. *Frontiers In Immunology*, *7*. doi: 10.3389/fimmu.2016.00388

Williams, R., Xue, B., Uversky, V., & Dunker, A. (2013). Distribution and cluster analysis of predicted intrinsically disordered protein Pfam domains. *Intrinsically Disordered Proteins*, *1*(1), e25724. https://doi.org/10.4161/idp.25724

World Health Organization. (2020). mRNA vaccines against COVID-19: Pfizer-BioNTech COVID-19 vaccine BNT162b2: prepared by the Strategic Advisory Group of Experts (SAGE) on immunization working group on COVID-19 vaccines, 22 December 2020. World Health Organization.

Yampolsky, L., & Stoltzfus, A. (2005). The Exchangeability of Amino Acids in Proteins. *Genetics*, *170*(4), 1459-1472. https://doi.org/10.1534/genetics.104.039107

Yang, Y., Li, F., Luo, X., Jia, B., Zhao, X., & Liu, B. et al. (2019). Identification of LCN1 as a Potential Biomarker for Breast Cancer by Bioinformatic Analysis. *DNA And Cell Biology*, *38*(10), 1088-1099. https://doi.org/10.1089/dna.2019.4843

Zhang, X., Zhang, Y., Ba, Z., Kyritsis, N., Casellas, R., & Alt, F. (2019). Fundamental roles of chromatin loop extrusion in antibody class switching. *Nature*, *575*(7782), 385-389. doi: 10.1038/s41586-019-1723-0

Wang, L. (2017). Early Diagnosis of Breast Cancer. *Sensors*, *17*(7), 1572. https://doi.org/10.3390/s17071572