n

Comparison of Epitopes and paratopes

Miruna Serian

2019

Abstract

This is the abstract

1 Introduction

2 Methods

2.1 Identifying Ig/Ag complexes

A preliminary data set of 72 experimentally determined Ig/Ag complexes associated with lysozyme where identified using IMGT's database. The antigens from the obtained complexes come from different species: Camelus dromedarius(14 complexes), shark - Ginglymostoma cirratum(4 complexes), mouse - Mus musculus (53 complexes). Three out of the complexes contain humanized antibodies.

2.2 Retrieving contacts

Annotation for the antibody and lysozyme chains for each pdb entry was obtained from the IMGT and was further used to collect information about the interactions between the antigen chain and both heavy and light chains from the antibody. The interacting residues' positions and their names were fetched from PDBSum and stored locally.

2.3 CDR extraction and filtering

A Multiple Sequence Alignment of the antibody structures present in the Ab-Ag complexes revealed that many of these complexes were redundant in the sense that they included numerous representatives of the same protein, for example as site-specific mutants.

The CDR sequences for both heavy and light chains for each antibody in the Ag/Ig complexes were extracted from PyIgClassify: a database of antibody CDR structural classifications using a script wrote in python. The CDRs sequences were compared was compared using similarity scores. The algorithm for generating the similarity scores uses an averaged string matching score computed for each pair of CDRs from two different Abs pair in the dataset. The results revealed that the heavy chain CDRs(42 unique) are more variable than the light chain CDRs (25 unique). For the purpose of this Ab/Ag interaction comparison, all but one Ag–Ab complex representative of a unique heavy chain CDR were deleted from the data set. The resulting data set is comprised of 38 Ab/Ag complexes associated with different species and their the corresponding coordinate files were collected from the Protein Data Bank both in pdb and fasta format using a python script based on pdb-tools.
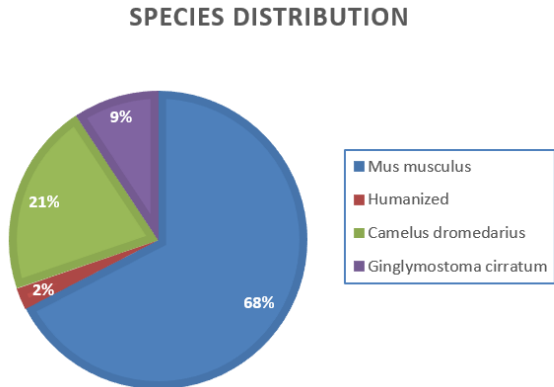
SPECIES DISTRIBUTION

Figure 1: Species Distribution of Ag with unique CDRs

2.4 Epitope identification

For the lysozyme chain in each complex, the amino acid sequences were downloaded using pdb-tools[] and were further aligned using EBI's Clustal Omega MSA tool. The generated alignment was downloaded and a python script was written to obtain the equivalent contact positions for every Ag sequence from the alignment, taking into account the existent insertions and gaps.

Epitopes were defined as the Ag residues that interact with Abs through both hydrogen bonds and non hydrogen bonds. The position of each contacting residue was stored together with the number of interactions with the Ab for further analyses. /bigbreak

To analyse the similarity between the lysozyme epitope in each of the complexes with unique Ab CDRs, the Jaccard index was computed for each pair of epitopes(antigen contact residues positions represented as numbers) and the average for each complex was plotted according to the following formula:

$$S_i = \frac{\sum_{j}^{N} J_{ij}}{N} \tag{1}$$

The Jaccard similarity coefficient is defined as the size of the intersection divided by the size of the union of two sets and can serve as an useful measure of similarity for two sets of data, with a range from 0 to 100.
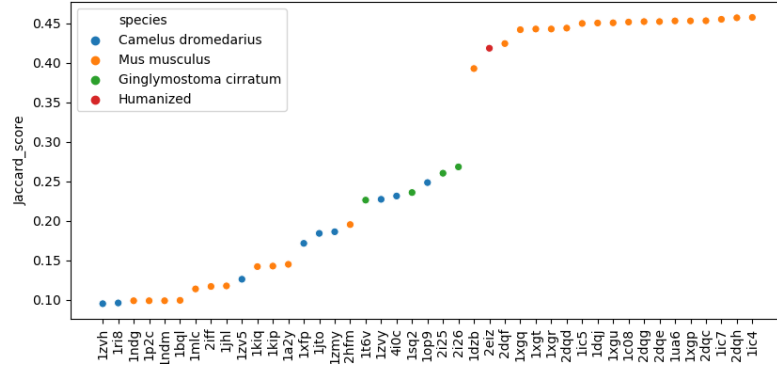
$$J_{ij} = \frac{(X \cap Y)}{(X \cup Y)}$$

Figure 2: Distribution of Jaccard Scores

Based on the distribution of the Jaccard scores a distinct group of similar epitopes in mouse can be observed. However, the grouping becomes less clear for the rest of the epitopes, partly due to the overrepresentation of the mouse complexes. The similar mouse subgroup was then removed from the data set and the Jaccard scores were computed again. The distribution of scores was used in conjunction with the map of contacting residues to further group the epitopes. The mouse group was split in two different groups with different sizes, of 17 epitopes and 3 respectively. Amongst the camel and shark epitopes, 3 subgroups of similarity were identified, two of them comprising 3 epitopes and one comprising 7 epitopes. The groups are presented in the Table 1.

| Group 1 | | Group 2 | | |
| Mus Musculus | | Camel and Shark | | |
Group 1a	Group 1b	Group 2a	Group 2b	Group 2c
2dqd	1bql	1jto	1zvy	1ri8
1xgr	1mlc	1jtp	1sq2	1rjc
1xgt	2iff	1jtt	1t6v	1zv5
1xgq		1xfp		
1dqj		1zmy		
2dqc		2i25		
1c08		2i26		
1xgu				
1ic5				
1ic4				
1ic7				
2dqh				
1ua6				
1xgp				
2dqe				
2dqg				

Figure 3: Distribution of Jaccard Scores

3 Results

3.1 Epitope binding

3.1.1 Amino Acid Composition of Epitopes

Hen egg white lysozyme is a small enzyme, comprised of a single polypeptide containing 129 amino acids. An initial Ab/Ag contacts analysis suggests that a large part of the residues in Ags interact with the Abs, with 65 % (84)of the residues making contact and 60%(78) making hydrogen bonds. At the level of amino acid composition, ASN, ARG and GLY are the most frequent residues found across the epitopes, in both H-bonded and non-H bonded contacts. In contrast, LEU, PHE, ILE, PRO, VAL are present in small numbers. This is in accordance with the general agreement that epitopes are enriched in charged and polar amino acids and depleted of aliphatic hydrophobic amino acids[4]. THR and TRP are the most frequent residues with a significantly higher number than the rest of the residues. This indicates it may play an important role in the recognition process. Only a moderate occurrence of the reported 'hotspot' binding residues TRP, ARG , TYR and ILE [1] is observed.
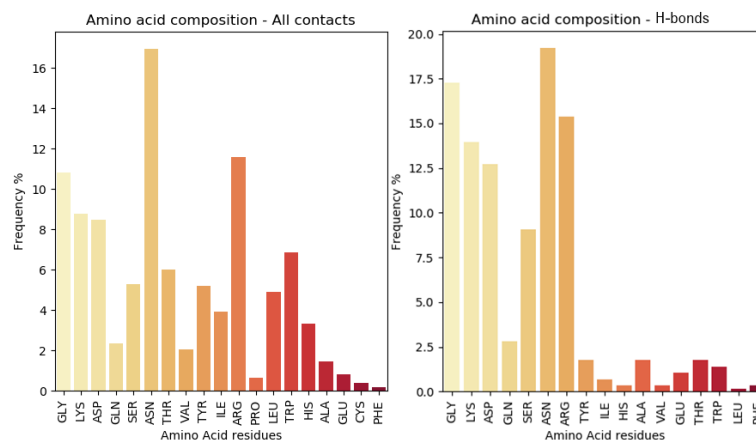
Figure 4: Distribution of amino acids on the epitope

3.1.2 Epitope Similarity

For each of the 5 groups, the H-bonded-only epitope residues and all bonded epitope residues were analysed. Two heatmaps for each group were generated to examine the epitope binding patterns. An overall epitope overlap is observed within all groups, with Group 1 having the most similar epitopes as well as being the largest group.

The number of hydrogen contacts is significantly lower than the number of non-H contacts in all groups. While the pattern of H-bonded contacts is more varied within groups, certain contacts appear to be highly conserved across members of groups. These interactions most likely strengthen the AB/Ag complex, while the aromatic residues that establish van der Waals and hydrophobic interactions drive the binding to the paratope[2]

Mouse associated epitopes

The epitopes of the first mouse group are comprised of over 25 residues, being the longest ones in the data set(Figure 5). The epitope contact residues represent a big proportion of the lysozyme surface area. It can be seen that certain epitope residues make considerably more contacts with the antigen compared to other residues such the ones are the equivalent positions 21(ARG), 97(Lys), 98(Lys) and 102(ASP). These residues are hydrophillic and are thought to contribute more significantly to complex stabilisation by being more energetically dominant [3]. What is more, these residues seem to be structurally close on the surface of the epitope and near the cleft of the lysozyme structure.
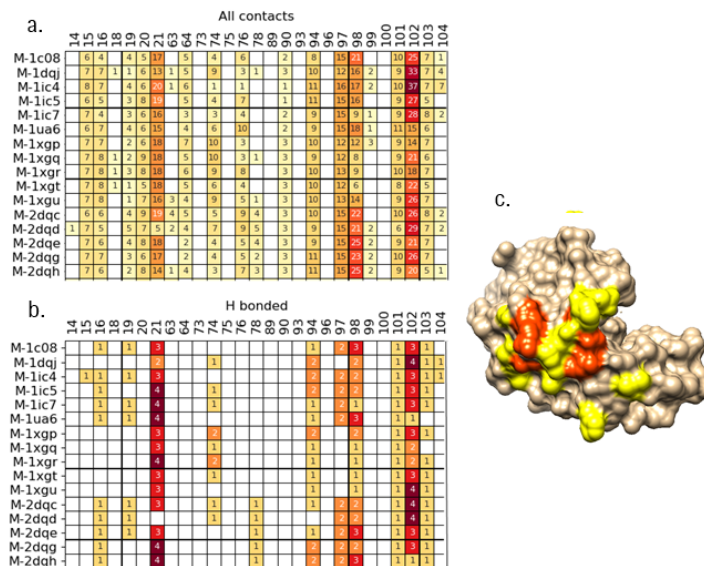
Figure 5: Group 1a - Number of contacts on epitope

The epitopes in the second mouse group 1b are composed of fewer residues, spanning the central area of the lysozyme sequence. The structural positions of these residue differ completely to the ones in the first mouse subgroup. At the structural level, they seem to be concentrated in one surface patch that is found in the other arm of the structure and on the opposite side of the structure compared to the first subgroup (Figure 6).
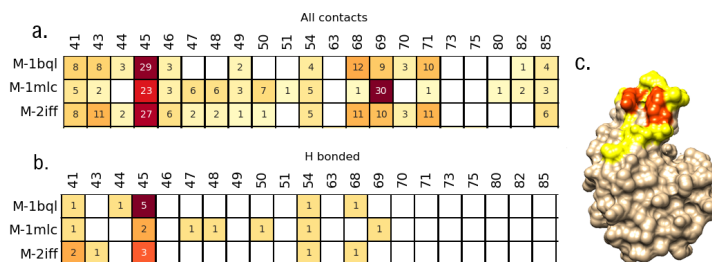

Figure 6: Group 1b - Number of contacts on epitope

Camel and shark associated epitopes

The epitopes show slight variation between the three groups of camel and shark associated complexes. Group 2b indicates the highest epitope residue conservation. While all the camel epitope residues in the group are part of the shark epitope, sharks seem to have an additional stretch of 6 residues at the beginning of the epitope sequence. In contrast to the mouse epitopes, the residues with the highest number of contacts are clustered in the pocket structure of the lysozyme(Figure 7 d.). Camelid heavy chain-only antibodies posses a unique mode of antigen recognition being able to interact with a vast repertoire of antigen binding sites. Furthermore, they can bind more hidden epitope residues and the clefts of antigens [2] [6].
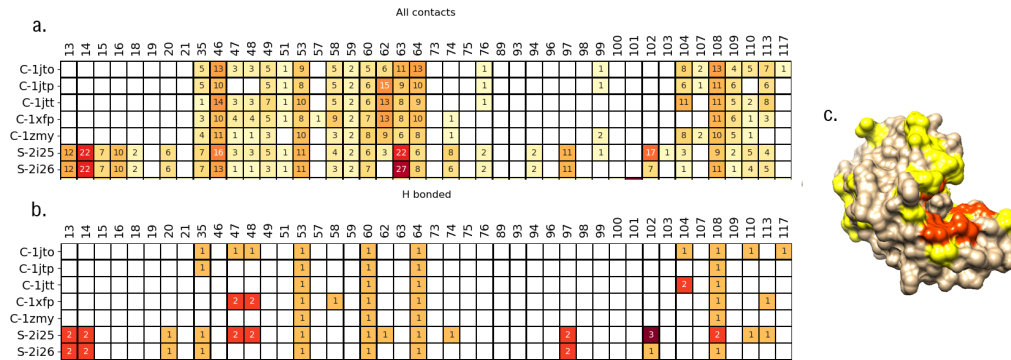
8

Figure 7: Group 2a - Number of contacts on epitope. Shark only residues in orange and common residues blue

The epitopes for subgroups 2b are smaller in size than the first subgroup and are found on opposite side on the lysozyme structure (Figure 8). Similarly to the other subgroup, the hydrogen bonded residues and the ones with higher number of contacts to the Ag are found in the pocket area of the structure. However, a smaller surface area interacts with the paratope through hydrogen bonds.

The third group epitopes are formed of residues at different positions compared to group 2b. The structural representation illustrates the positioning of the H-bonded epitope residues on the opposite side of the antigen.
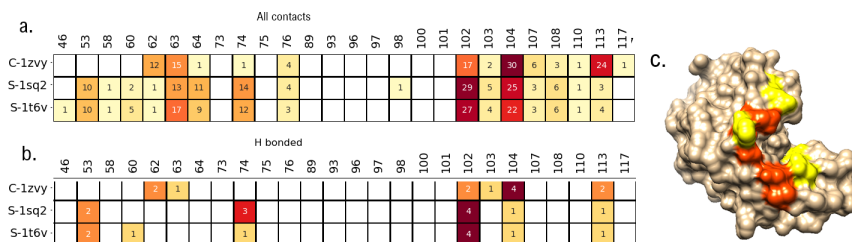

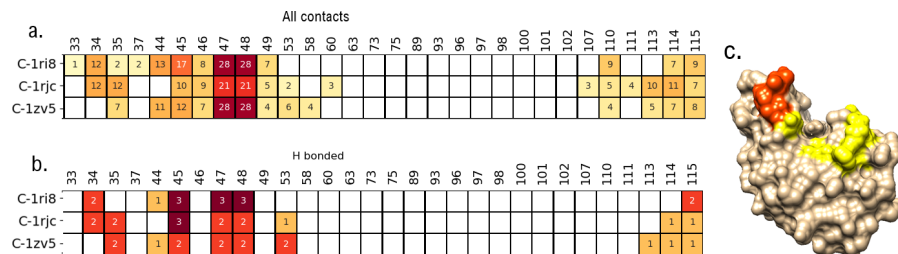Figure 8: Group 2b - Number of contacts on epitope

Figure 9: Group 2c - Number of contacts on epitope

A number of the residues are also observed to be conserved not only within groups but also across the groups of different species. Such residues include residue 64(Gr 1a, Gr 2a, Gr 2b), 76(Gr 1b, Gr 2a, Gr 2b) and 102(Gr 1a, Gr2a, Gr 2b).

9

3.2 Paratope binding

3.2.1 Antibody sequence variation

For the comparison of the binding pattern of different paratopes to the same epitope, the complexes containing Abs with similar sequences were removed. As the interacting residues usually reside in the Fab region of the antibody, the Abs were filtered out based on their CDR sequences, removing the complexes with identical heavy chains. Heavy chains were found to be more varied than the light chains. To study the overall similarity of the antibodies, a program was written in python to compute pairwise global alignments and generate a similarity score using eq.

$$Sim_i = \frac{\sum_{j}^{N} S_{ij}}{N} \tag{2}$$

,where Sim_i is the similarity score of sequence i, S_{ij} is the score of the global alignment of sequences i and j, textitN is the number of sequences in the group

3.2.2 Paratope amino acid composition

The paratope amino acid composition for each species varies slightly. A big presence of TYR is observed in all groups, while the frequency of the other amino acids is different within species. Consistent with previous reports on lysozyme, a great amount of TYR and ASN is found in all species. TYR plays an especially important role in interactions with epitopes having the highest number of contacts with the paratope [5]. An unusually high number of GLY and ARG residues are present in the shark paratopes. Gly in CDR-H3 was found to play an important role in the conformation of the antigen for increased binding affinity, as reported by a study on antibody library design [3]. The highest diversity of residue types is found in sharks epitopes which, unlike in mouse and camel, contain LEU, LYS, ALA and VAL in small numbers. Interesting to note is the shark epitopes contain a higher than the average of TYR, ASN and ser residues, a moderate number of ASP and GLN and very small numbers of GLU, GLY and ARG. Mouse epitopes follow the same pattern but with more moderate numbers of SER, ASN and TYR. This result suggests that, the variability of the antibody paratope residues is more limited in camel and mouse, and that shark antibodies are potentially able to bind a greater variety of antibodies.

Previous work suggests that amino acid preference of epitopes and paratopes differs from the general composition of the antigen and antibody surface (Andersen et al., 2006; Ofran et al., 2008;
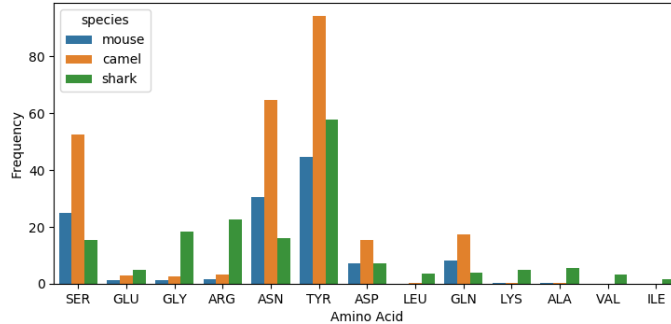

Figure 10: Paratope Amino acid Occurrence

Because some of the residues in the epitope were observed to bind more than one position in the antibody. To analyse the pattern of binding as seen in figure **??**

3.2.3 CDR binding

To investigate the binding modes of the different Ags to similar epitopes, the heavy chain CDR binding preference was presented in the form of heatmaps. The peripheral epitope residues bind almost exclusively H2 while most of the central residues interact with all three CDRs. Exceptions to this case include residue 101 which is conserved within the group and binds H2 and H3 and residue 99 which only binds H1. Complexes 1xgbp, 1xgq, 1xgt and 1xgu show a loss of CDR binding preference at residues 64, 74 and 76.
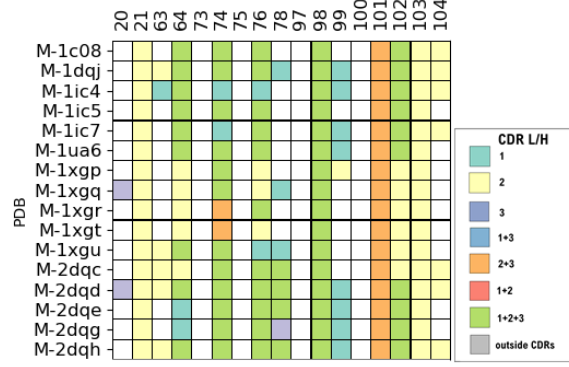

Figure 11: mouse - CDR binding preference

In the smaller second mouse subgroup, complexes 1bql and 2iff have an almost identical binding pattern, while 1mlc suggests some minor differences Out of the epitope residues that interact with the heavy chain through hydrogen bonding, half of them bind outside the CDRs and usually their binding preference is conserved within groups.

The camel and shark subgroups reveal an interesting distribution of epitope binding, with only a small percentage of residues binding within the CDRs. The hydrogen-bonded contacts almost entirely bind outside the CDRs.
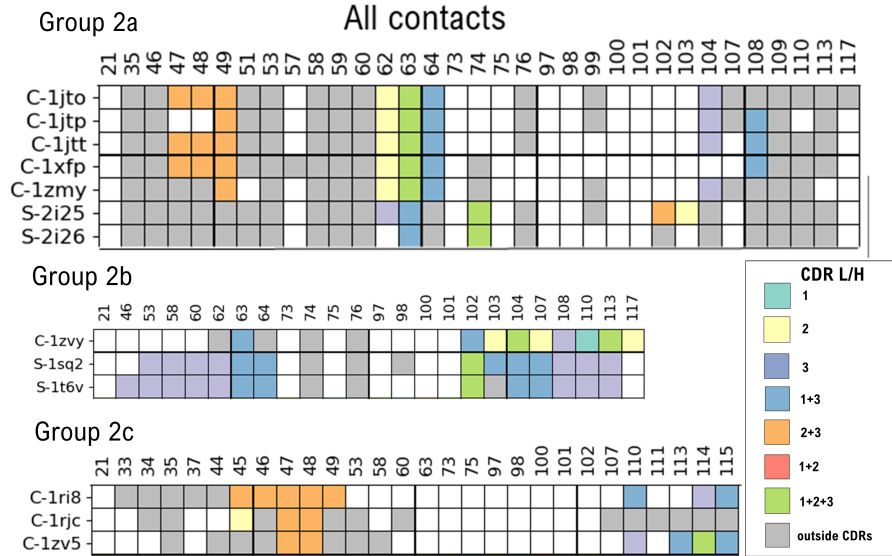

Figure 12: Camel and Shark - CDR binding preference

The structures of the antibodies do don vary greatly within groups. Although the type of the amino acids of each Heavy chain CDR that interacts with the epitope seems to vary, the position of these residues in the structures is conserved across the members of the groups(Fig 13- only Group 2b H3 residues that bind the epitope are coloured in red and Fig 14 where Group 1b interacting H1 residues are coloured cyan, H2 residue in yellow, H3 residues in purple)

3.2.4 Framework region binding

A further investigation of only the epitope residues binding outside the heavy chain CDRs suggests a similar pattern preference for the different FR(Framework Regions) in the mouse subgroups.
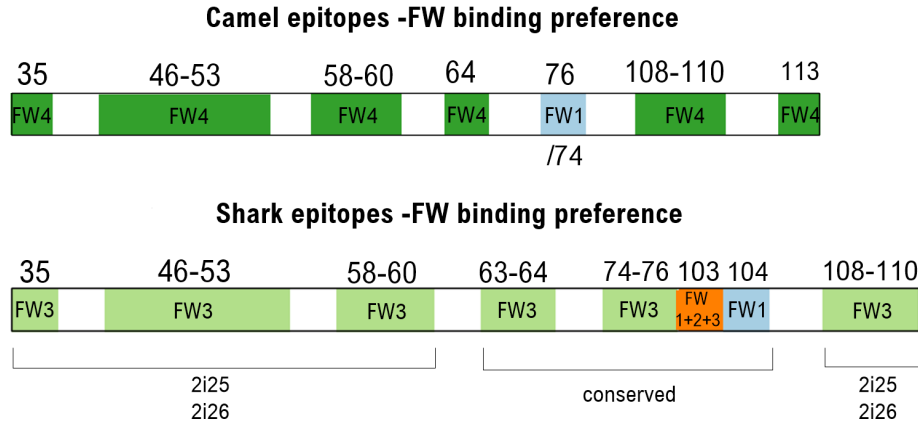

Figure 13: Camel and Shark - CDR binding preference

For the camel and shark complexes there is no distinguishable binding pattern. A great proportion of the shark epitopes bind FR3 and a small number bind both FR1 and FR2. On the other hand, the camel associated epitopes do not appear to bind FR3, and mostly bind FR1 and FR3. Interesting to note is the fact that none of the heavy chain epitopes studied here bind FR4 and that FR1 does not make any individual contact with the epitopes,but only in conjunction with FR3.

At the light chain level, the epitopes associated with mouse antibodies bind in a very small proportion outside the CDRs, and only bind FR2.

3.2.5 Amino acid composition of Heavy chain CDR contact residues

The amino acid composition of the heavy chain residues were coloured using Shapely colouring scheme. In mouse Group 1a there is no great variation between the type of residues across the members

[I will probably delete the consensus] Some of the residues in the epitope were observed to bind more than one position in the antibody. To analyse the pattern of binding, a consensus residue was determined for each position based on the occurrence of residues. Further, for each contacting residue was then assigned a score of -1, 0, 1 according to the following criteria:

- -1 if the antigen residue is different to consensus

- 0 if the epitope residue is binding to multiple residues on the antigen, including the consensus

- 1 if the epitope residue only binds the consensus

References

[1] Andrew A Bogan and Kurt S Thorn. Anatomy of hot spots in protein interfaces. *Journal of Molecular Biology*, 280(1):1–9, July 1998.

[2] Aline Desmyter, Thomas R. Transue, Mehdi Arbabi Ghahroudi, Minh-Hoa Dao Thi, Freddy Poortmans, Raymond Hamers, Serge Muyldermans, and Lode Wyns. Crystal structure of a camel single-domain VH antibody fragment in complex with lysozyme. *Nature Structural & Molecular Biology*, 3(9):803–811, September 1996.

[3] Frederic A. Fellouse, Kaori Esaki, Sara Birtalan, Demetrios Raptis, Vincenzo J. Cancasci, Akiko Koide, Parkash Jhurani, Mark Vasser, Christian Wiesmann, Anthony A. Kossiakoff, Shohei Koide, and Sachdev S. Sidhu. High-throughput generation of synthetic antibodies from highly functional minimalist phage-displayed libraries. *Journal of Molecular Biology*, 373(4):924–940, November 2007.

[4] Jens Vindahl Kringelum, Morten Nielsen, Søren Berg Padkjær, and Ole Lund. Structural analysis of b-cell epitopes in antibody:protein complexes. *Molecular Immunology*, 53(1-2):24–34, 2013.

[5] Minh N Nguyen, Mohan R Pradhan, Chandra Verma, and Pingyu Zhong. The interfacial character of antibody paratopes: analysis of antibody–antigen structures. *Bioinformatics*, 33(19):2971–2976, June 2017.

[6] R. L. Stanfield. Crystal structure of a shark single-domain antibody v region in complex with lysozyme. *Science*, 305(5691):1770–1773, September 2004.