EPITOPIC

Application Note

Positional Scanning of Antibody Epitopes and Peptide Ligands

Date: Sept 17th 2024

Statistical positional scanning of epitopes is essential for a deep analysis of the large amount of statistical data obtained in epitope fingerprinting. This improves the handling of the large number of sequences sharing epitope motifs. With no extra experimental efforts or consumption of materials it is generating data comparable to the positional scanning using synthetic peptides.

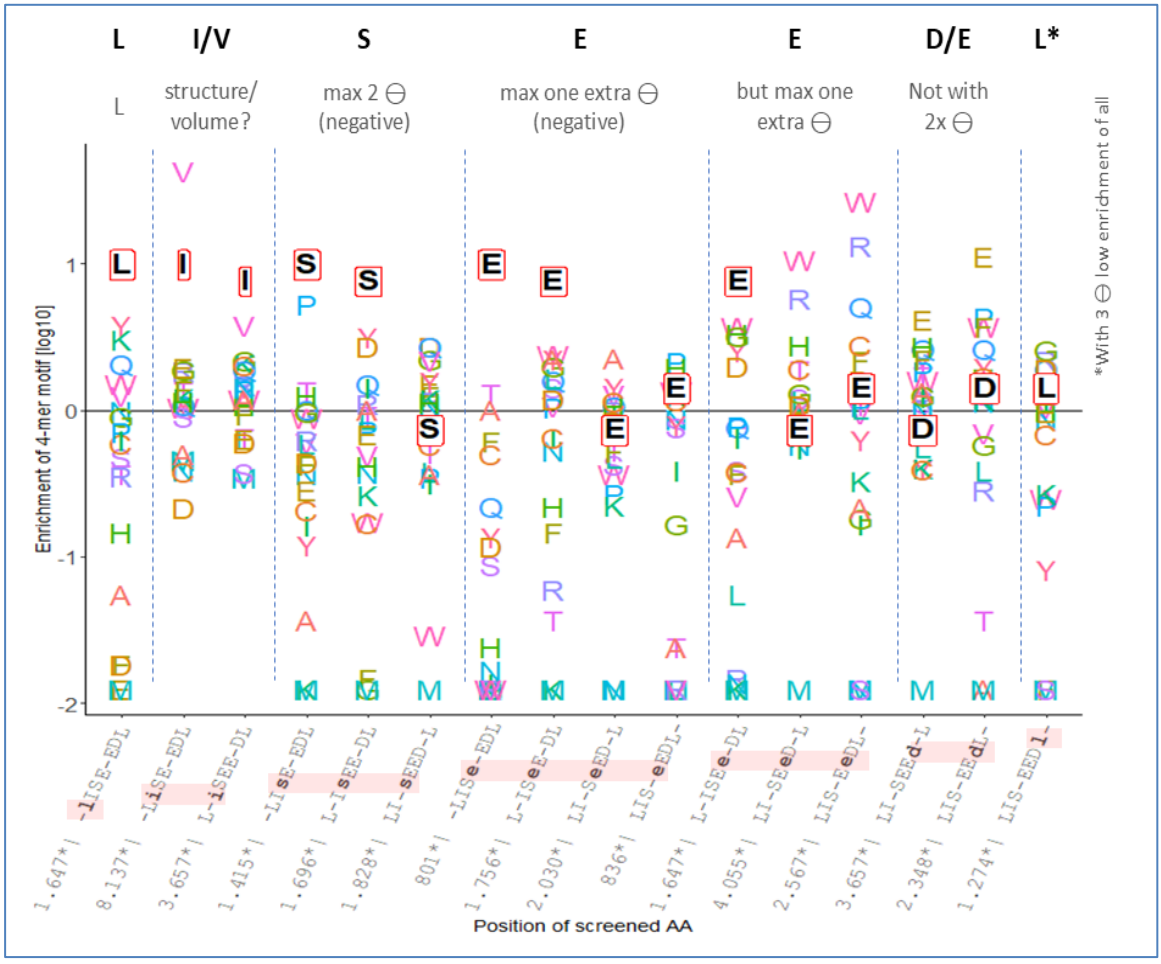


Figure 1: Statistical positional scanning of the epitope of the 9E10 (anti-c-myc) antibody

epitopic's unique statistical approach to identify peptide ligands and to map epitopes of monoclonal antibodies offers a shortcut to understanding the relevant residues of hundreds or thousands of peptide variations bound by a target. A classical approach of positional scanning is expandable when making use of complex data obtained from a single set of experiments with a unique peptide library. epitopic's library is a naïve collection of 16-mer peptides that are random, but their statistical composition is well controlled.

Interactions between two proteins can often be reduced to a few key interactions. The importance of each individual amino acid contributing to an epitope also depends on neighboring amino acid residues. This is true for both synthetic peptides and mutational studies of proteins. Changes in both binding and non-binding residues can alter the structure and interactions within a peptide epitope, or generate steric hindrance for binding to an antibody, not to mention additive effects that would require the generation of N-dimensional arrays of peptides for N positions. These complex interactions within protein epitopes highlight the need for sophisticated analytical methods to fully understand and predict antibody-antigen binding. Computational approaches, such as molecular dynamics simulations and machine learning algorithms, are increasingly employed to model these intricate relationships and predict the effects of amino acid substitutions. Additionally, combinatorial screening techniques such as phage display and yeast surface display have become valuable tools for mapping epitopes and identifying critical residues involved in antibody recognition in complex molecules. The ultimate proof may be considered a crystal structure; however, even here, individual amino acid side chains are affected by the packing of molecules in the crystal. In addition to the cost and time involved in these studies, the outcome often requires confirmation through further analyses.

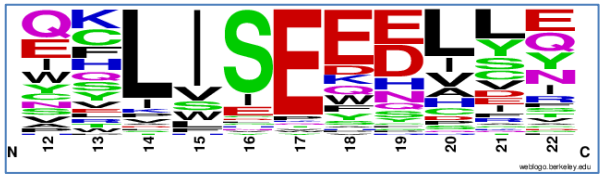
These advanced screening techniques provide a wealth of information about antibody-antigen interactions, but they often require substantial time and resources. epitopic's epitope fingerprinting offers a more efficient alternative, allowing for rapid analysis of billions of peptide variants. This high-throughput approach cannot only complement traditional methods, but also change the process of epitope mapping and antibody characterization in various research and therapeutic applications. epitopic's epitope fingerprinting allows data to be accessed for the statistical analysis of thousands of peptide variants preselected by binding to an antibody or another target molecule. Therefore, these data should only contain valuable information. This reduces the inevitable amount of no-sense information generated by combinatorial approaches used in the positional scanning of multiple amino acid residues.

Example: Monoclonal antibody 9E10 (anti-c-myc peptide)

The monoclonal antibody 9E10 was raised against a peptide of the c-myc protein and has been used successfully in many different applications for decades. Its epitope has been studied using all available methods described in the literature. We mapped the epitope using our statistical approach (see report on the epitopic home page[1]). Here, we provide a more detailed analysis that resembles the more recent efforts using amino acid substitutions in synthetic peptides. However, the data were obtained from a single set of experiments that enriched the peptides from the epitopic library.

The alignment on the right of sequences found at least 12x and with 5 amino acid identity with the antigen sequence already reveals the essentials of the epitope.

More than 12,000 different enriched sequences in the screened data contain parts of the antigen sequence. According to an alignment of 1,500 of these the epitope is:



Examples partially compiled from recent reports:

EQKLISEEdL, X-ray structure[2]
eqKLISEedLn, peptide filter array[3]
X(Q,E)XLISEXX(L,M), consensus from 116 selected sequences, mRNA display[4]
qKL(I/V)SEedl, peptide array. incl. positional scanning[5]

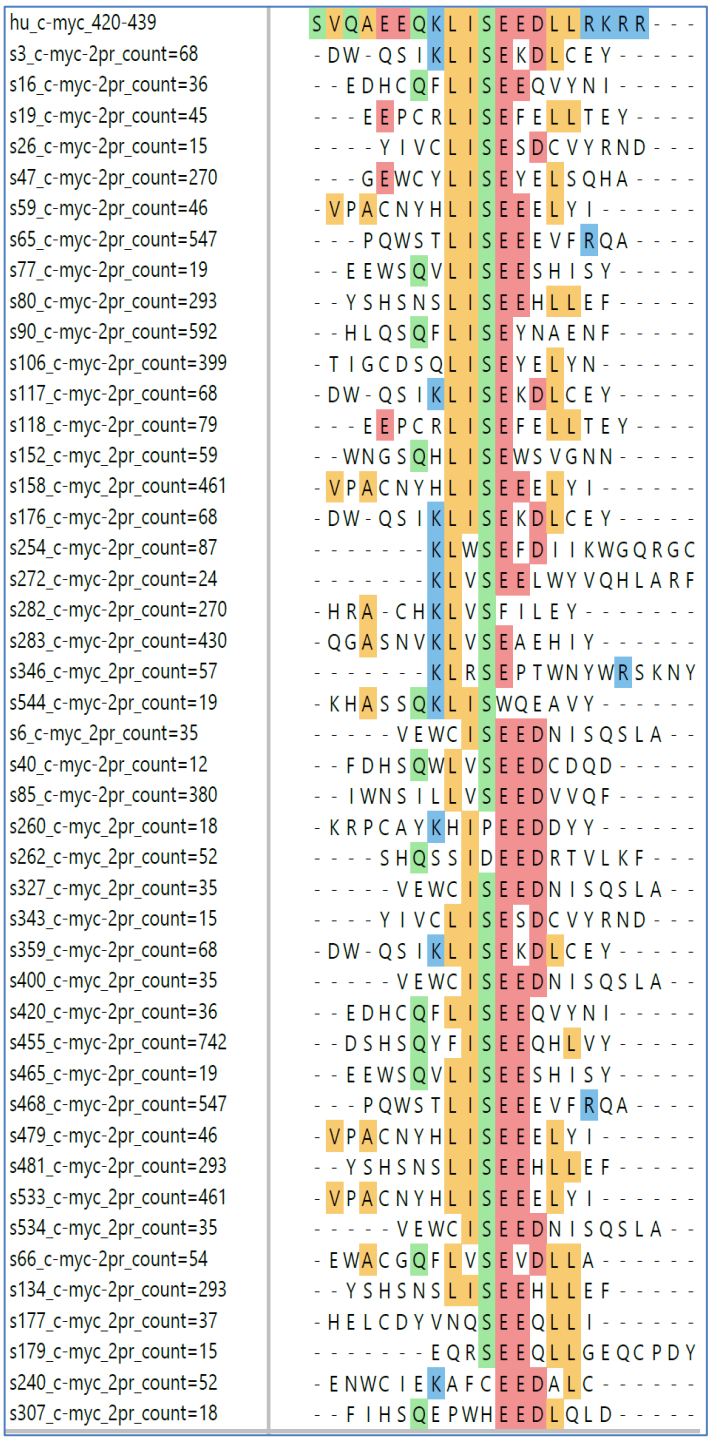


Figure 2: Selected sequences found at least 12x with at least 5 positions identity with the antigen.

[1] https://www.epitopic.com/downloads.html
[2] Krauss et al. 2008; doi: 10.1002/prot.22080
[3] Hilpert et.al. 2001; doi:10.1093/protein/14.10.
[4] Baggio et. al. 2002; https://doi.org/10.1002/jmr.567
[5] https://www.pepperprint.com/fileadmin/downloads/Application_Notes/Application_Note_Validation_of_a_Recombinant_Human_anti-Myc_Antibody.pdf

Going deeper into the 9E10 epitope with statistical positional scanning

All methods, including the alignment of sequences with similarities to the antigen, have the disadvantage of neither expecting nor having the capacity to analyze variations in neighboring sequences. Physical binding measurement methods often cover only to 1-2 orders of magnitude of binding variance. epitopic's approach statistically covers up to three orders of difference in enrichment and can therefore be considered superior. However, one can only conclude that some preferences are given for the number of negative charges, (ir-)relevance of the side chain, and so on. There are simply too many sequences for a quick visual or even standard alignment analysis, such as the weblogo[6].

True statistical positional analysis by enrichment can be performed by running all variations of the 4-mer motifs to build the potential epitope. If we use data from the naïve library, no clear preference is observed. The probability of amino acids in each position of a 4-mer is unbiased resp. predictable within a narrow range of variations.

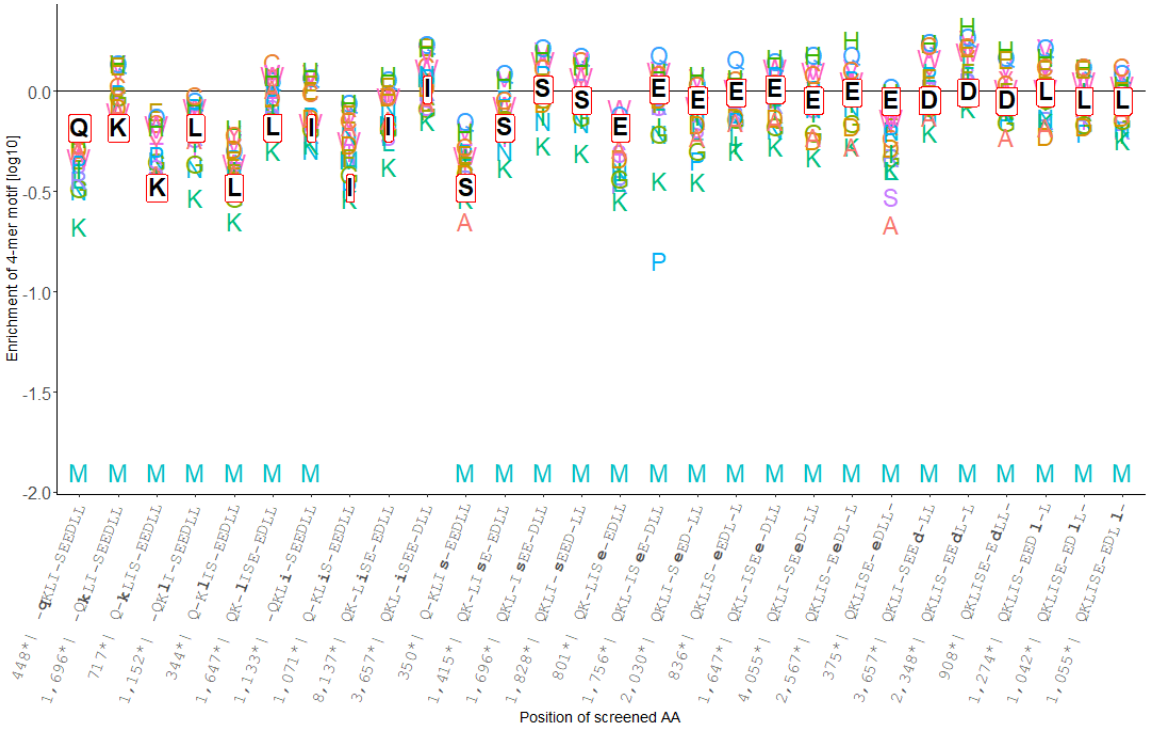


*Figure 3: Naive library data set (2.1 Mio sequences) screened for LISEEDL 4-mers with the enrichment of all amino acids versus library design and covering all four positions, i.e. *AAA, A*AA, AA*A, AAA*. X-axis is log10 enrichment versus theoretical value, which is close to no deviation at all. The data is based on a total of ten times more sequences than in the next figure 4.*

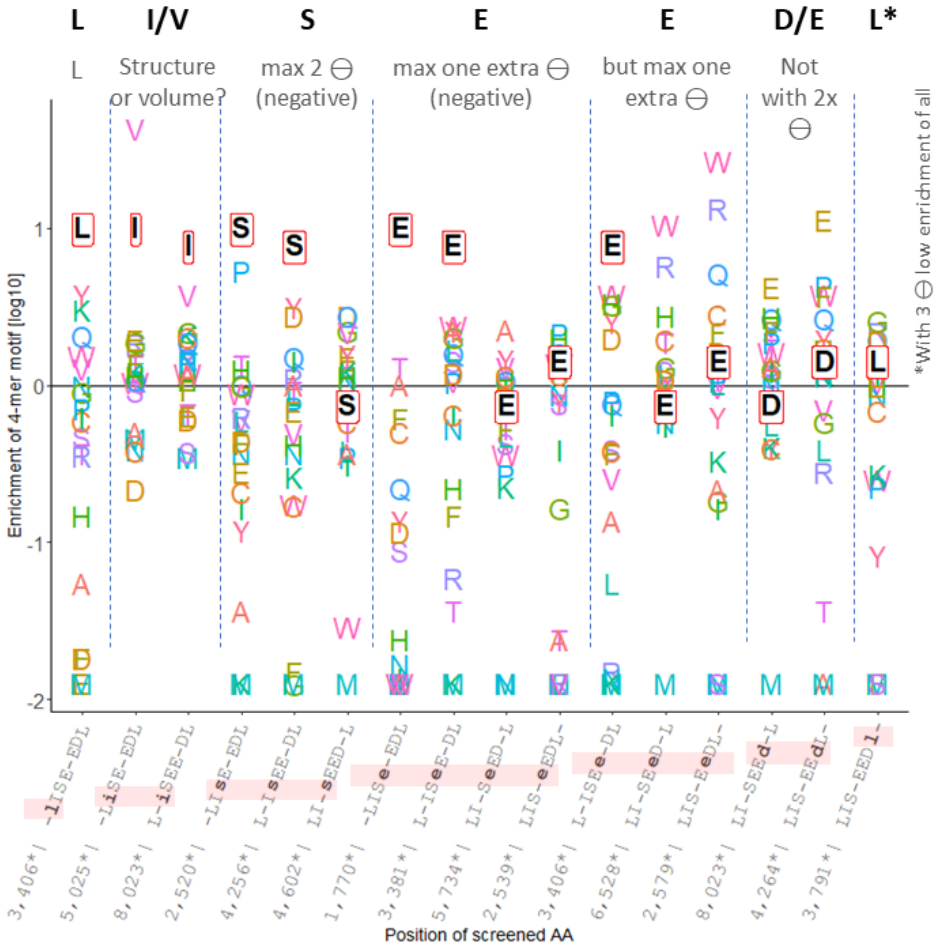
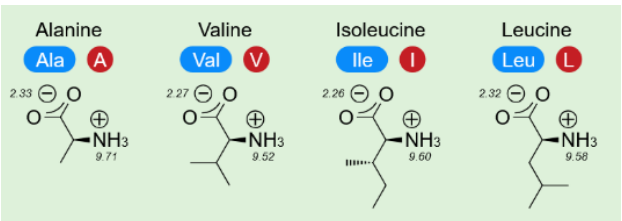


Figure 4: NGS data (750.000 sequences) from library selected on the 9E10 antibody screened for LISEEDL 4-mers. See text and figure 3 for explanations.

For the statistical positional scanning in figure 4 we used data from a selection of the antibody bound to immunotubes. In the starting library all 4-mers are expected to be present at approximately the same frequency, which can also be calculated using the design of the random library. Each row presents the log10 value of the enrichment for all variations in one position in a given 4-mer of the epitope. On the left the enrichment is calculated for *ISE, i.e. "lISE-EDL." The next positions are "LiSE-EDL" and "L-iSEE-DL," where it becomes obvious that Val is highly preferred in the first context but less in the second. This indicates that the antibody-binding site probably accommodates a branched amino acid such as Val or Ile, but none of the related amino acids.[6]



[6] From TungstenEinsteinium - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=117005436

The question is, of course: How relevant are these figures of 10-100-fold enrichment. The number of sequences in the NGS dataset with the ISE 3-mer analyzed is given in the X-axis description. In addition, with respect to the numbers in other positions, the L(I/V)SE 4-mers were enriched significantly based on a sufficiently large number of individual sequences. Please compare also figure 3 and figure 4 considering that the naïve data set contained many more sequences than the data set obtained after selection.

The main results of this screening are shown at the top of the graph. As already might be concluded from the alignment, there were only a few sequences with three negative charges. Since the bound peptides have a length of 16 amino acids, the reason may not be the negative charge influence on the peptide, but the binding process may be more favorable with only two negative charges. This is also supported by the fact that Glu is preferred over Asp at the second last position analyzed.

Digging deeper into the data can further reveal, that in the last position bulky amino acids are less preferred than Leu, although with "LIS-EEDl" we are looking at a unfavorable and less enriched motif with three negative charges.

Interestingly, the extension of this analysis to EQKLISEEDLL provides no preferred amino acids. Apparently at these edges numerous other variations in the 16-mer library peptide can compensate for "loss" of these amino acids usually included in descriptions of the epitope.

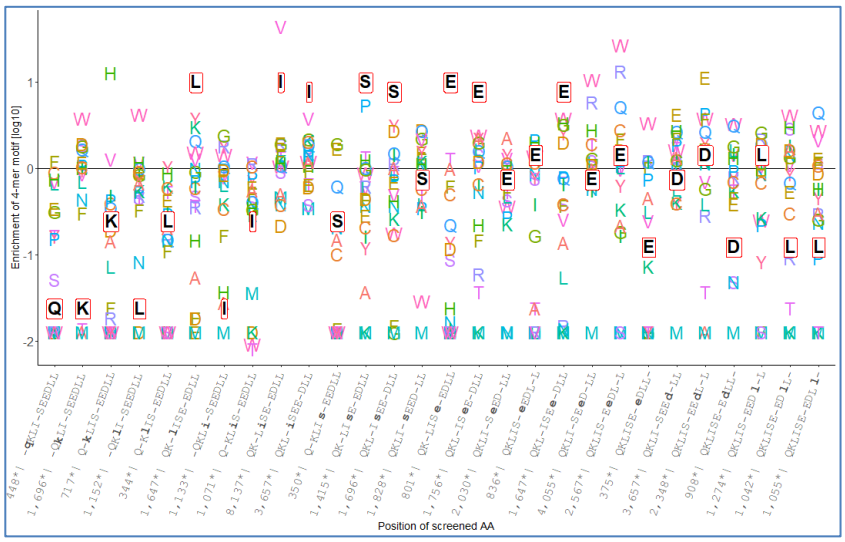


Figure 5: After selection on a target the enrichment of non-relevant motifs is either unchanged or even reduced.

Conclusion: Statistical positional scanning of epitopes is essential for a deep analysis of the large amount of statistical data obtained in epitope fingerprinting. This improves the handling of the large number of sequences that share epitope motifs. With no extra experimental effort or consumption of materials, it generates data comparable to positional scanning using synthetic peptides.

Acknowledgement: This application note was generated using data and input from Dr. Nicolas Delaroque and Dr. Karolin Kern from the Fraunhofer IZI in Leipzig, Germany.