EPITOPIC

Epitope Fingerprinting Report

Date: September 29th, 2017

Target: anti-muCD184 BD 551966

Antigen: mu CD184

Panning conditions: Immobilized on M270 carboxy beads

General comments

Epitope fingerprinting of anti-muCD184 on M270 carboxy beads was carried out. This report summarizes the results of identifying the epitope.

Panning and NGS library statistics

Table 1: NGS data. "1pr" refers to the first and "2pr" refers to the second round of selection. Seq Count: Number of sequences in MiSeq raw data. Valid Count: Number of sequences with no detectable sequencing error. 3-mer / 4-mer motifs: The number of motifs in the first round should be close to theory, i.e. ca 7200 resp. 132000 for medium size data sets of up to 500000 sequences. Panning: Protocol used.

Database

Panning	Data Set Name	Seq Count	Valid Count	3-mer motifs	4-mer motifs
Beads	anti-muCD184-1pr	212 846	142 381	7 254	131 316
Beads	anti-muCD184-2pr	357 662	244 407	7 254	129 907

Statistical Analyses

Data based on all selection rounds' 4-mer motifs in the data sets plotted against mu CD184 sequence. The reference data set (red) contains almost 2 Mio Sequences from the naïve library.

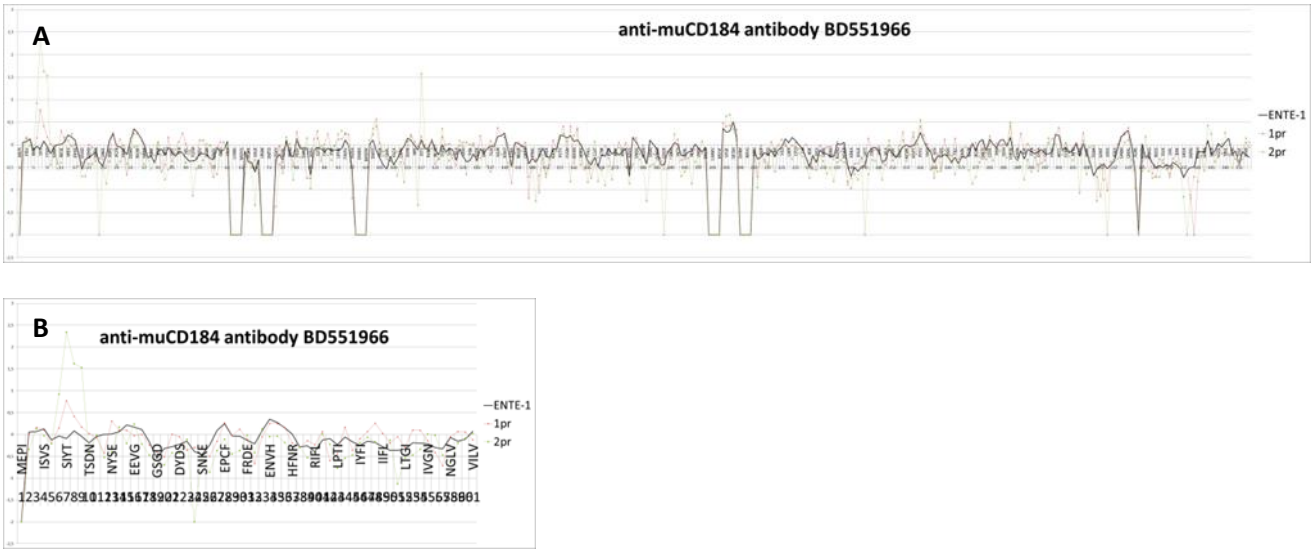

Figure 1: Statistic of 4-mer motifs. X-axis: All possible 4mer fragments of the antigen starting from 1-4, 2-5 ...to... end. Y-axis: log of enrichment vs. calculated probability in the data set, e.g. 3 is 10^3 resp. 1000-fold enrichment. Control ENTE1 naïve starting library has max 0.5 deviation from 0 (black). After the first panning round (red) the library is depleted of motifs, many approach a cut off of <-2, which should not be taken as a save value with respect to the number of sequences. Because this would mean that the motif is less than 1/100 of the expected average, which usually is less than 1 sequence in a million.

Epitope

Most likely epitope sequence is:

7-(S)IYTSDnyxeE-17

According to database entries the Ser is not present in the mature protein, the C-terminal Glu is only revealed by the alignment.

This epitope is based on the statistical analysis based on the protein sequence in Fig. 4. Genetic variants may lead to other results.

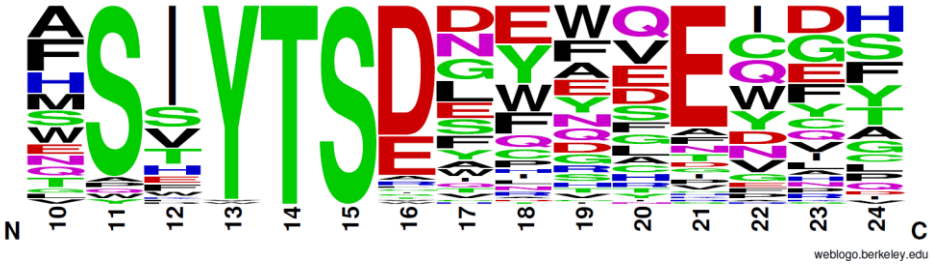

Figure 2: Presentation as web logo. Fingerprint in Web-Logo style from all sequences containing at least 4 Amino acids of IYTSD from 2nd selection round. The web-logo is available online: http://weblogo.berkeley.edu/

Fingerprint

The Weblogo is based on each sequence counted once, the following table for the fingerprint counts each found sequence, i.e. enriched sequences with potentially higher affinity have stronger weight. This shows that certain positions have only preferences for certain features of the amino acids, as for example the second Tyr may be replaced by Phe and Trp. Results from the >1550 sequences carrying motif YTSD from the second selection round.

Antigen				S	I	YTSD	N	Y	S	E	E
	A5	A4	A3	A2	A1	M	B1	B2	B3	B4	B5
C	0	2	0	1	1	YTSD	0	13	6	1	0
P	1	3	0	5	3	YTSD	1	16	1	0	2
G	1	0	2	0	1	YTSD	248	0	1	29	25
A	0	0	5	3	0	YTSD	10	2	61	8	1
V	8	9	1	0	418	YTSD	6	0	13	355	7
I	2	54	0	1	275	YTSD	48	5	0	1	0
L	1	1	7	0	7	YTSD	7	7	1	13	0
M	0	0	20	0	0	YTSD	0	0	0	0	0
F	7	0	6	0	12	YTSD	743	365	15	1	1
Y	4	1	0	3	0	YTSD	2	253	1	1	0
W	3	4	0	1	6	YTSD	25	824	248	0	0
T	0	7	0	0	748	YTSD	11	2	4	2	2
S	0	0	1	100	24	YTSD	15	0	6	244	0
N	48	2	6	0	1	YTSD	6	0	736	0	2
Q	7	9	9	2	0	YTSD	4	3	11	860	0
H	9	5	58	0	13	YTSD	8	0	351	3	0
K	0	2	0	0	2	YTSD	1	0	0	0	1
R	9	3	0	0	0	YTSD	1	0	4	7	0
E	0	9	2	1	58	YTSD	2	67	84	14	1504
D	0	6	0	0	1	YTSD	421	0	10	14	8
	100	117	117	117	1570		1559	1557	1553	1553	1553

Figure 3: Presentation as EPITOPIC fingerprint. Fingerprint of the motif YTSD containing 1559 sequences of the second selection round. Amino acid positions framed in black are part of the antigen sequence.

Epitope in sequence

```
>sp|P70658|CXCR4_MOUSE C-X-C chemokine receptor type 4 OS=Mus musculus OX=10090
GN=Cxcr4 PE=1 SV=2
MEPISVSIYTSDNYSEEVGSGDYDSNKEPCFRDENVHFNRIFLPTIYFIIFLTGIVGNGL
VILVMGYQKKLRSMTDKYRLHLSVADLLFVITLPFWAVDAMADWYFGKFLCKAVHIIYTV
NLYSSVLILAFISLDRYLAIVHATNSQRPRKLLAEKAVYVGVWIPALLLTIPDFIFADVS
QGDISQGDDRYICDRLYPDSLWMVVFQFQHIMVGLILPGIVILSCYCIIISKLSHSKGHQ
KRKALKTTVILILAFFACWLPYYVGISIDSFILLGVIKQGCDFESIVHKWISITEALAFF
HCCLNPILYAFLGAKFKSSAQHALNSMSRGSSLKILSKGKRGGHSSVSTESESSSFHSS
```

Figure 4: Epitope in protein sequence of mu CD184. The sequence labelled in red is the most likely epitope of the antibody.

Figure 5: Sequence Examples. Alignment of enriched sequences from the second selection round with the enriched 4-mer motifs reveals a complex motif. This alignment lists only sequences found more than 2x.

Epitope in 3D-Structure

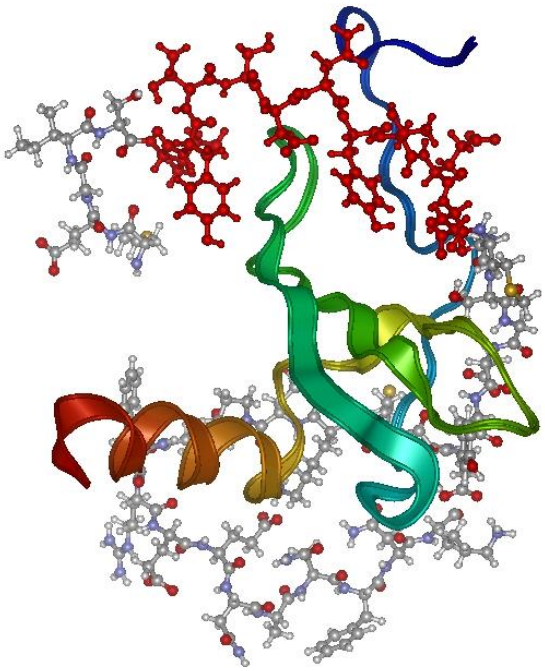

Figure 6: 3D-structure of mu CD184. The epitope is highlighted in red in the NMR structure 2K04.pdb. This structure shows the N-terminus of the CXCR4 domain in complex with stromal cell-derived factor 1 (shown as ribbon).

The rather extended structure fits well to the statistically observed fingerprint. This structure also contains the enriched N-terminal Ser of the enriched sequences.

Summary

The data presented here shows strong evidence that the epitope sequence of the antibody is **(S)IYTSDnyxeE**.