Epitope Fingerprinting Report

Date: October 23rd, 2024

Work Package: Epitope Fingerprinting for one monoclonal antibody

Task: Epitope identification for anti-mouse CD184 (BD551966, Rat IgG2b)

General comments

The CD184 antibody was subjected to epitope fingerprinting. The epitope was determined from using NGS data of peptide phage of the naïve ENTE-1 library selected on the antibody coupled to M270 Carboxy magnetic beads (EDC/NHS coupling).

A sequence pool of 142,381 resp. 244,407 sequences from a first and a second selection round was analysed for statistical enrichment of motifs corresponding to potential epitopes in the protein. An excellent enrichment of sequences corresponding to a discontinuous epitope was found.

This report is based on data analysed earlier. In the last years epitopic has improved many statistical analysis tools. The analysis of the data is now not only faster, but we are also able to produce even more details of the antibodies' binding to the antigen.

In an internal challenge we managed to run the analysis with more details revealed in a much shorter time. With the excellent datasets it took only about eight hours for the entire work presented here.

The results show/explain:

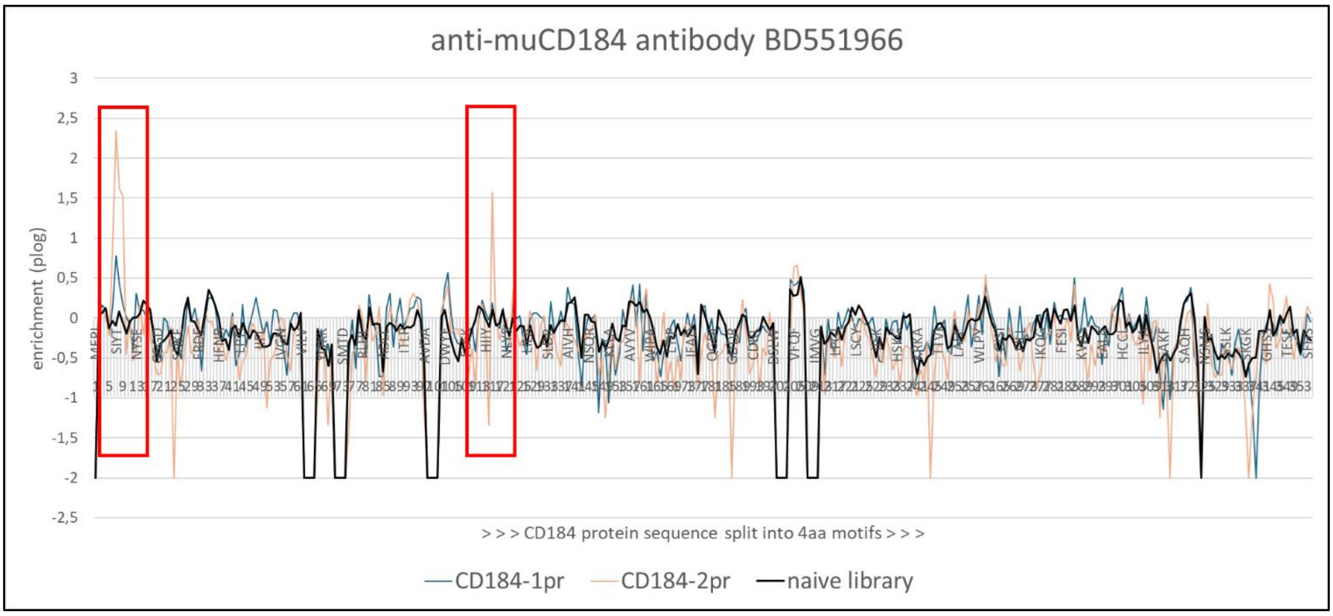
- **Epitope details by amino acids recognized**
- **Structural confirmation**
- **Species specificity**

Statistical Analyses

Standard analysis based on the frequency of 4-mer motifs of the antigen sequence resulted in an overview of potential epitopes.

The graphic below shows the statistical enrichment of the consecutive 4-mer peptides of the CD184 protein sequence in the data sets after one and two selection rounds. The Y-axis shows the enrichment of 4-mers over the theoretically expected value defined by the library design. The scale is log10, so several motifs occur hundredfold and more frequently in the NGS datasets than expected. These expected values can be calculated, because the naïve library is constructed from just one codon per amino acid and behaves stable even upon replication. The black line shows the actually measured statistical data from >2 Mio. sequences of the naïve library. Deviations are due to variations in the coupling yield of codon building blocks during the synthesis.

The boxes indicate motifs within the protein enriched by more than hundredfold. Such peaks could belong to positions in the protein bound by the antibody, and the sequences containing them were further analysed.



This "enrichment" curve shows that these positions represent the motifs:
6-VSIYTSDN-13, 117-IYTV-120
It can be seen below that the second motif's frequency is due to an overall enrichment of sequences containing IYTx, where x can represent a wide variety of amino acids.

The following step in the analysis is the retrieval of full-length sequences from the selected library pool, which are containing enriched motifs. These are usually hundreds of similar sequences. They are aligned with and without the antigen sequence. In most cases the alignments were at least cured for sequences found in the datasets only once, twice or even more frequently. If a sequence is found at least three to five times, it is less likely to contain typical NGS sequencing errors.

Alignment of sequences with enriched motifs

All sequences containing the amino acid 4-mers covering the partial CD184 sequence

5-SVSIYTSDNYSEEVGSG-21

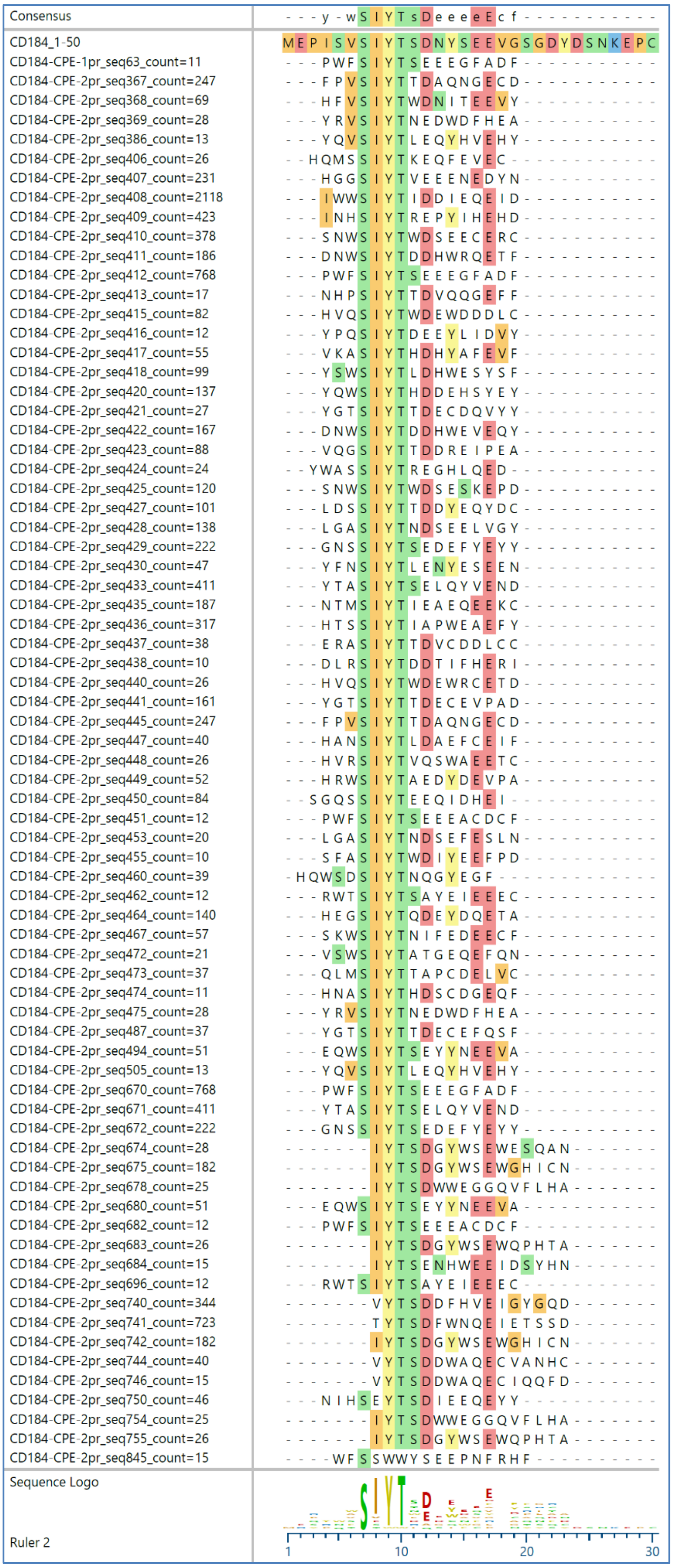
were retrieved from the two selection's NGS datasets. This identified 905 different in total 13,068 sequences. To keep the number of sequences displayed in the alignment to a minimum, the alignment displays only sequences found at least tenfold and sharing at least 5 amino acid identities with the aligned antigen sequence.

The epitope fingerprinting avoids multiple selection rounds. Therefore, hundreds of similar sequences can be used for the analysis of the spectrum of peptides recognized by the antibody. Due to the ability of statistical analysis of a balanced library design the selected 13,000 sequences are making up only a fraction of the 400,000 sequences of the analysed data sets.

The huge number of sequences in the ENTE-1 library (ca 5x10^9) allows also to identify multiple sequences sharing more than eight amino acid identity with the antigen. The probability to identify a **single** sequence with eight identities by chance is roughly 4x10^{-11} (=1/20^8) or 2x10^{-5} in this dataset of 500,000 sequences.

Based on the alignment and considering the amino acid residues shared by the enriched sequences the epitope could be best described as

sIYTxDxyxeE



```
Consensus                          ---y-wSIYTsDeeeeEcf-----------
CD184_1-50                         MEPISVSIYTSDNYSEEVGSGDYDSNKEPC
CD184-CPE-1pr_seq63_count=11       ---PWFSIYTSEEEGFADF-----------
CD184-CPE-2pr_seq367_count=247     ---FPVSIYTTDAQNGECD-----------
CD184-CPE-2pr_seq368_count=69      ---HFVSIYTWDNITEEVY-----------
CD184-CPE-2pr_seq369_count=28      ---YRVSIYTNEDWDFHEA-----------
CD184-CPE-2pr_seq386_count=13      ---YQVSIYTLEQYHVEHY-----------
CD184-CPE-2pr_seq406_count=26      --HQMSSIYTKEQFEVEC------------
CD184-CPE-2pr_seq407_count=231     ---HGGSIYTVEEENEDYN-----------
CD184-CPE-2pr_seq408_count=2118    ---IWWSIYTIDDIEQEID-----------
CD184-CPE-2pr_seq409_count=423     ---INHSIYTREPYIHEHD-----------
CD184-CPE-2pr_seq410_count=378     ---SNWSIYTWDSEECERC-----------
CD184-CPE-2pr_seq411_count=186     ---DNWSIYTDDHWRQETF-----------
CD184-CPE-2pr_seq412_count=768     ---PWFSIYTSEEEGFADF-----------
CD184-CPE-2pr_seq413_count=17      ---NHPSIYTTDVQQGEFF-----------
CD184-CPE-2pr_seq415_count=82      ---HVQSIYTWDEWDDDLC-----------
CD184-CPE-2pr_seq416_count=12      ---YPQSIYTDEEYLIDVY-----------
CD184-CPE-2pr_seq417_count=55      ---VKASIYTHDHYAFEVF-----------
CD184-CPE-2pr_seq418_count=99      ---YSWSIYTLDHWESYSF-----------
CD184-CPE-2pr_seq420_count=137     ---YQWSIYTHDDEHSYEY-----------
CD184-CPE-2pr_seq421_count=27      ---YGTSIYTTDECDQVYY-----------
CD184-CPE-2pr_seq422_count=167     ---DNWSIYTDDHWEVEQY-----------
CD184-CPE-2pr_seq423_count=88      ---VQGSIYTTDDREIPEA-----------
CD184-CPE-2pr_seq424_count=24      --YWASSIYTREGHLQED------------
CD184-CPE-2pr_seq425_count=120     ---SNWSIYTWDSESKEPD-----------
CD184-CPE-2pr_seq427_count=101     ---LDSSIYTTDDYEQYDC-----------
CD184-CPE-2pr_seq428_count=138     ---LGASIYTNDSEELVGY-----------
CD184-CPE-2pr_seq429_count=222     ---GNSSIYTSEDEFYEYY-----------
CD184-CPE-2pr_seq430_count=47      ---YFNSIYTLENYESEEN-----------
CD184-CPE-2pr_seq433_count=411     ---YTASIYTSELQYVEND-----------
CD184-CPE-2pr_seq435_count=187     ---NTMSIYTIEAEQEEKC-----------
CD184-CPE-2pr_seq436_count=317     ---HTSSIYTIAPWEAEFY-----------
CD184-CPE-2pr_seq437_count=38      ---ERASIYTTDVCDDLCC-----------
CD184-CPE-2pr_seq438_count=10      ---DLRSIYTDDTIFHERI-----------
CD184-CPE-2pr_seq440_count=26      ---HVQSIYTWDEWRCETD-----------
CD184-CPE-2pr_seq441_count=161     ---YGTSIYTTDECEVPAD-----------
CD184-CPE-2pr_seq445_count=247     ---FPVSIYTTDAQNGECD-----------
CD184-CPE-2pr_seq447_count=40      ---HANSIYTLDAEFCEIF-----------
CD184-CPE-2pr_seq448_count=26      ---HVRSIYTVQSWAEETC-----------
CD184-CPE-2pr_seq449_count=52      ---HRWSIYTAEDYDEVPA-----------
CD184-CPE-2pr_seq450_count=84      --SGQSSIYTEEQIDHEI------------
CD184-CPE-2pr_seq451_count=12      ---PWFSIYTSEEEACDCF-----------
CD184-CPE-2pr_seq453_count=20      ---LGASIYTNDSEFESLN-----------
CD184-CPE-2pr_seq455_count=10      ---SFASIYTWDIYEEFPD-----------
CD184-CPE-2pr_seq460_count=39      -HQWSDSIYTNQGYEGF-------------
CD184-CPE-2pr_seq462_count=12      ---RWTSIYTSAYEIEEEC-----------
CD184-CPE-2pr_seq464_count=140     ---HEGSIYTQDEYDQETA-----------
CD184-CPE-2pr_seq467_count=57      ---SKWSIYTNIFEDEECF-----------
CD184-CPE-2pr_seq472_count=21      ---VSWSIYTATGEQEFQN-----------
CD184-CPE-2pr_seq473_count=37      ---QLMSIYTTAPCDELVC-----------
CD184-CPE-2pr_seq474_count=11      ---HNASIYTHDSCDGEQF-----------
CD184-CPE-2pr_seq475_count=28      ---YRVSIYTNEDWDFHEA-----------
CD184-CPE-2pr_seq487_count=37      ---YGTSIYTTDECEFQSF-----------
CD184-CPE-2pr_seq494_count=51      ---EQWSIYTSEYYNEEVA-----------
CD184-CPE-2pr_seq505_count=13      ---YQVSIYTLEQYHVEHY-----------
CD184-CPE-2pr_seq670_count=768     ---PWFSIYTSEEEGFADF-----------
CD184-CPE-2pr_seq671_count=411     ---YTASIYTSELQYVEND-----------
CD184-CPE-2pr_seq672_count=222     ---GNSSIYTSEDEFYEYY-----------
CD184-CPE-2pr_seq674_count=28      -------IYTSDGYWSEWESQAN-------
CD184-CPE-2pr_seq675_count=182     -------IYTSDGYWSEWGHICN-------
CD184-CPE-2pr_seq678_count=25      -------IYTSDWWEGGQVFLHA-------
CD184-CPE-2pr_seq680_count=51      ---EQWSIYTSEYYNEEVA-----------
CD184-CPE-2pr_seq682_count=12      ---PWFSIYTSEEEACDCF-----------
CD184-CPE-2pr_seq683_count=26      -------IYTSDGYWSEWQPHTA-------
CD184-CPE-2pr_seq684_count=15      -------IYTSENHWEEIDSYHN-------
CD184-CPE-2pr_seq696_count=12      ---RWTSIYTSAYEIEEEC-----------
CD184-CPE-2pr_seq740_count=344     -------VYTSDDFHVEIGYGQD-------
CD184-CPE-2pr_seq741_count=723     -------TYTSDFWNQEIETSSD-------
CD184-CPE-2pr_seq742_count=182     -------IYTSDGYWSEWGHICN-------
CD184-CPE-2pr_seq744_count=40      -------VYTSDDWAQECVANHC-------
CD184-CPE-2pr_seq746_count=15      -------VYTSDDWAQECIQQFD-------
CD184-CPE-2pr_seq750_count=46      ---NIHSEYTSDIEEQEYY-----------
CD184-CPE-2pr_seq754_count=25      -------IYTSDWWEGGQVFLHA-------
CD184-CPE-2pr_seq755_count=26      -------IYTSDGYWSEWQPHTA-------
CD184-CPE-2pr_seq845_count=15      ----WFSSWWYSEEPNFRHF----------
Sequence Logo
Ruler 2                            1        10        20        30
```

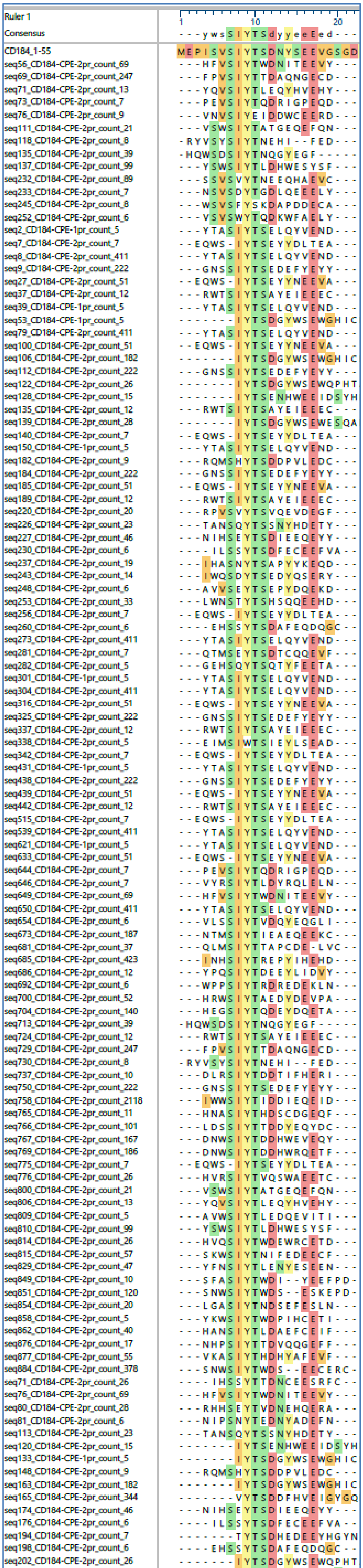
The amino acid sequence of the epitope is not a continuous stretch. Therefore, it makes sense to search for discontinuous motifs, too. The following result is obtained when sequences are retrieved from the data base containing 5-mer variations of the type AxAAA, AAxAA, AAAxA of the motifs `SVSIY`, `SIYTS`, `YTSDN`, `SDNYS` or `NYSEE` (taken from 5-`SVSIYTSDNYSEEVGSG`-21). Here x can be any amino acid.

This approach can possibly identify more sequences by allowing variations of amino acids instead of strict 4-mer motifs. In this case, it expands the alignment above by only showing the sequences which are found at least 5 times and that have 5 or more amino acids identity to the antigen sequence. But the sequences in total only repeat the pattern which was visible in the previous alignment already.

Based on the second alignment the epitope would be best described as below. Here for example xYx can be replaced by one or two other hydrophobic amino acids and a high probability for two Glu being recognized by the antibody.

vsIYTSdxYxEEv

In general such amino acids must not always be interacting with the antibody, but they might also be essential for a proper fold of the peptide. One example could be the Asp in the center of the sequence.



```
Ruler 1                              1        10        20
Consensus                            ---ywsSIYTSdyyeeEed---
CD184_1-55                           MEPISVSIYTSDNYSEEVGSGD
seq56_CD184-CPE-2pr_count_69         ---HFVSIYTWDNITEEVY---
seq69_CD184-CPE-2pr_count_247        ---FPVSIYTTDAQNGECD---
seq71_CD184-CPE-2pr_count_13         ---YQVSIYTLEQYHVEHY---
seq73_CD184-CPE-2pr_count_7          ---PEVSIYTQDRIGPEQD---
seq76_CD184-CPE-2pr_count_9          ---VNVSIYEIDDWCEERD---
seq111_CD184-CPE-2pr_count_21        ---VSWSIYTATGEQEFQN---
seq118_CD184-CPE-2pr_count_8         -RYVSYSIYTNEHI--FED---
seq135_CD184-CPE-2pr_count_39        -HQWSDSIYTNQGYEGF-----
seq137_CD184-CPE-2pr_count_99        ---YSWSIYTLDHWESYSF---
seq232_CD184-CPE-2pr_count_89        ---SSVSVYTNEEQHAEVC---
seq233_CD184-CPE-2pr_count_7         ---NSVSDYTGDLQEEELY---
seq245_CD184-CPE-2pr_count_8         ---WSVSFYSKDAPDDECA---
seq252_CD184-CPE-2pr_count_6         ---VSVSWYTQDKWFAELY---
seq2_CD184-CPE-1pr_count_5           ---YTASIYTSELQYVEND---
seq7_CD184-CPE-2pr_count_7           --EQWS-IYTSEYYDLTEA---
seq8_CD184-CPE-2pr_count_411         ---YTASIYTSELQYVEND---
seq9_CD184-CPE-2pr_count_222         ---GNSSIYTSEDEFYEYY---
seq27_CD184-CPE-2pr_count_51         --EQWS-IYTSEYYNEEVA---
seq37_CD184-CPE-2pr_count_12         ---RWTSIYTSAYEIEEEC---
seq39_CD184-CPE-1pr_count_5          ---YTASIYTSELQYVEND---
seq53_CD184-CPE-1pr_count_5          -------IYTSDGYWSEWGHIC
seq79_CD184-CPE-2pr_count_411        ---YTASIYTSELQYVEND---
seq100_CD184-CPE-2pr_count_51        --EQWS-IYTSEYYNEEVA---
seq106_CD184-CPE-2pr_count_182       -------IYTSDGYWSEWGHIC
seq112_CD184-CPE-2pr_count_222       ---GNSSIYTSEDEFYEYY---
seq122_CD184-CPE-2pr_count_26        -------IYTSDGYWSEWQPHT
seq128_CD184-CPE-2pr_count_15        -------IYTSENHWEEIDSYH
seq135_CD184-CPE-2pr_count_12        ---RWTSIYTSAYEIEEEC---
seq139_CD184-CPE-2pr_count_28        -------IYTSDGYWSEWESQA
seq140_CD184-CPE-2pr_count_7         --EQWS-IYTSEYYDLTEA---
seq150_CD184-CPE-1pr_count_5         ---YTASIYTSELQYVEND---
seq182_CD184-CPE-2pr_count_9         ---RQMSHYTSDDPVLEDC---
seq184_CD184-CPE-2pr_count_222       ---GNSSIYTSEDEFYEYY---
seq185_CD184-CPE-2pr_count_51        --EQWS-IYTSEYYNEEVA---
seq189_CD184-CPE-2pr_count_12        ---RWTSIYTSAYEIEEEC---
seq220_CD184-CPE-2pr_count_20        ---RPVSVYTSVQEVDEGF---
seq226_CD184-CPE-2pr_count_23        ---TANSQYTSSNYHDETY---
seq227_CD184-CPE-2pr_count_46        ---NIHSEYTSDIEEQEYY---
seq230_CD184-CPE-2pr_count_6         ----ILSSYTSDFECEEFVA--
seq237_CD184-CPE-2pr_count_19        ---IHASNYTSAPYYKEQD---
seq243_CD184-CPE-2pr_count_14        ---IWQSDYTSEDYQSERY---
seq248_CD184-CPE-2pr_count_6         ---AVVSEYTSEPYDQEKD---
seq253_CD184-CPE-2pr_count_33        ---LWNSTYTSHSQQEEHD---
seq256_CD184-CPE-2pr_count_7         --EQWS-IYTSEYYDLTEA---
seq260_CD184-CPE-2pr_count_6         ----EHSSYTSDAFEQDQGC--
seq273_CD184-CPE-2pr_count_411       ---YTASIYTSELQYVEND---
seq281_CD184-CPE-2pr_count_7         ---QTMSEYTSDTCQQEVF---
seq282_CD184-CPE-2pr_count_5         ---GEHSQYTSQTYFEETA---
seq301_CD184-CPE-1pr_count_5         ---YTASIYTSELQYVEND---
seq304_CD184-CPE-2pr_count_411       ---YTASIYTSELQYVEND---
seq316_CD184-CPE-2pr_count_51        --EQWS-IYTSEYYNEEVA---
seq325_CD184-CPE-2pr_count_222       ---GNSSIYTSEDEFYEYY---
seq337_CD184-CPE-2pr_count_12        ---RWTSIYTSAYEIEEEC---
seq338_CD184-CPE-2pr_count_5         ---EIMSIWTSIEYLSEAD---
seq342_CD184-CPE-2pr_count_7         --EQWS-IYTSEYYDLTEA---
seq431_CD184-CPE-1pr_count_5         ---YTASIYTSELQYVEND---
seq438_CD184-CPE-2pr_count_222       ---GNSSIYTSEDEFYEYY---
seq439_CD184-CPE-2pr_count_51        --EQWS-IYTSEYYNEEVA---
seq442_CD184-CPE-2pr_count_12        ---RWTSIYTSAYEIEEEC---
seq515_CD184-CPE-2pr_count_7         --EQWS-IYTSEYYDLTEA---
seq539_CD184-CPE-2pr_count_411       ---YTASIYTSELQYVEND---
seq621_CD184-CPE-1pr_count_5         ---YTASIYTSELQYVEND---
seq633_CD184-CPE-2pr_count_51        --EQWS-IYTSEYYNEEVA---
seq644_CD184-CPE-2pr_count_7         ---PEVSIYTQDRIGPEQD---
seq646_CD184-CPE-2pr_count_7         ---VYRSIYTLDYRQLELN---
seq649_CD184-CPE-2pr_count_69        ---HFVSIYTWDNITEEVY---
seq650_CD184-CPE-2pr_count_411       ---YTASIYTSELQYVEND---
seq654_CD184-CPE-2pr_count_6         ---VLSSIYTVDQYEQGLI---
seq673_CD184-CPE-2pr_count_187       ---NTMSIYTIEAEQEEKC---
seq681_CD184-CPE-2pr_count_37        ---QLMSIYTTAPCDE-LVC--
seq685_CD184-CPE-2pr_count_423       ---INHSIYTREPYIHEHD---
seq686_CD184-CPE-2pr_count_12        ---YPQSIYTDEEYLIDVY---
seq692_CD184-CPE-2pr_count_6         ---WPPSIYTRDREDEKLN---
seq700_CD184-CPE-2pr_count_52        ---HRWSIYTAEDYDEVPA---
seq704_CD184-CPE-2pr_count_140       ---HEGSIYTQDEYDQETA---
seq713_CD184-CPE-2pr_count_39        -HQWSDSIYTNQGYEGF-----
seq724_CD184-CPE-2pr_count_12        ---RWTSIYTSAYEIEEEC---
seq729_CD184-CPE-2pr_count_247       ---FPVSIYTTDAQNGECD---
seq730_CD184-CPE-2pr_count_8         -RYVSYSIYTNEHI--FED---
seq737_CD184-CPE-2pr_count_10        ---DLRSIYTDDTIFHERI---
seq750_CD184-CPE-2pr_count_222       ---GNSSIYTSEDEFYEYY---
seq758_CD184-CPE-2pr_count_2118      ---IWWSIYTIDDIEQEID---
seq765_CD184-CPE-2pr_count_11        ---HNASIYTHDSCDGEQF---
seq766_CD184-CPE-2pr_count_101       ---LDSSIYTTDDYEQYDC---
seq767_CD184-CPE-2pr_count_167       ---DNWSIYTDDHWEVEQY---
seq769_CD184-CPE-2pr_count_186       ---DNWSIYTDDHWRQETF---
seq775_CD184-CPE-2pr_count_7         --EQWS-IYTSEYYDLTEA---
seq776_CD184-CPE-2pr_count_26        ---HVRSIYTVQSWAEETC---
seq800_CD184-CPE-2pr_count_21        ---VSWSIYTATGEQEFQN---
seq806_CD184-CPE-2pr_count_13        ---YQVSIYTLEQYHVEHY---
seq809_CD184-CPE-2pr_count_5         ---AVWSIYTLEDQEVITI---
seq810_CD184-CPE-2pr_count_99        ---YSWSIYTLDHWESYSF---
seq814_CD184-CPE-2pr_count_26        ---HVQSIYTWDEWRCETD---
seq815_CD184-CPE-2pr_count_57        ---SKWSIYTNIFEDEECF---
seq829_CD184-CPE-2pr_count_47        ---YFNSIYTLENYESEEN---
seq849_CD184-CPE-2pr_count_10        ---SFASIYTWDI--YEEFPD-
seq851_CD184-CPE-2pr_count_120       ---SNWSIYTWDS--ESKEPD-
seq854_CD184-CPE-2pr_count_20        ---LGASIYTNDSEFESLN---
seq858_CD184-CPE-2pr_count_5         ---YKWSIYTWDPIHCETI---
seq862_CD184-CPE-2pr_count_40        ---HANSIYTLDAEFCEIF---
seq876_CD184-CPE-2pr_count_17        ---NHPSIYTTDVQQGEFF---
seq877_CD184-CPE-2pr_count_55        ---VKASIYTHDHYAFEVF---
seq884_CD184-CPE-2pr_count_378       ---SNWSIYTWDS--EECERC-
seq71_CD184-CPE-2pr_count_26         ----IHSSYTTDNCEESRFC--
seq76_CD184-CPE-2pr_count_69         ---HFVSIYTWDNITEEVY---
seq80_CD184-CPE-2pr_count_28         ---RHHSEYTVDNEHQERA---
seq81_CD184-CPE-2pr_count_6          ---NIPSNYTEDNYADEFN---
seq113_CD184-CPE-2pr_count_23        ---TANSQYTSSNYHDETY---
seq120_CD184-CPE-2pr_count_15        -------IYTSENHWEEIDSYH
seq133_CD184-CPE-1pr_count_5         -------IYTSDGYWSEWGHIC
seq148_CD184-CPE-2pr_count_9         ---RQMSHYTSDDPVLEDC---
seq163_CD184-CPE-2pr_count_182       -------IYTSDGYWSEWGHIC
seq165_CD184-CPE-2pr_count_344       -------VYTSDDFHVEIGYGQ
seq174_CD184-CPE-2pr_count_46        ---NIHSEYTSDIEEQEYY---
seq176_CD184-CPE-2pr_count_6         ----ILSSYTSDFECEEFVA--
seq194_CD184-CPE-2pr_count_7         -------TYTSDHEDEEYHGYN
seq198_CD184-CPE-2pr_count_6         ----EHSSYTSDAFEQDQGC--
seq202_CD184-CPE-2pr_count_26        -------IYTSDGYWSEWQPHT
```

CD184 epitope location in the protein sequence

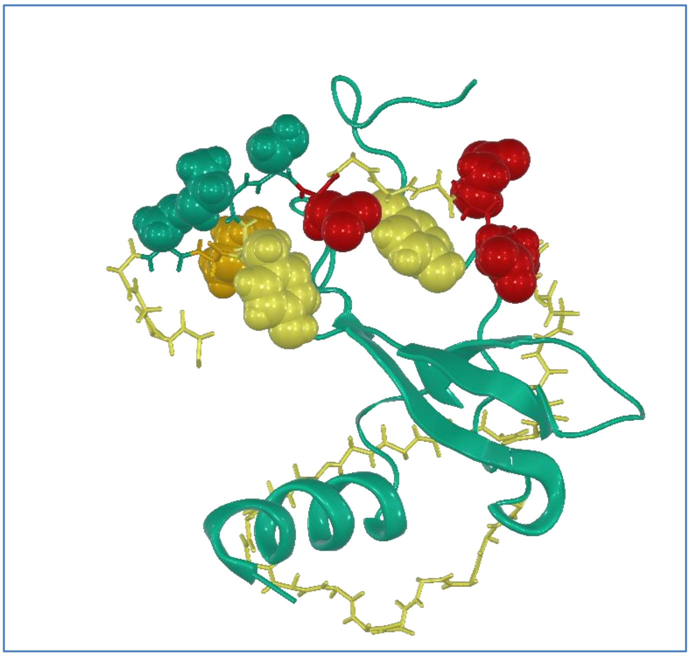
This is an overview of the full-length CD184 sequence where those regions are marked in yellow, that show enrichment in the statistical analysis of 4-mer motifs in the NGS datasets.

```
> CXCR4_MOUSE C-X-C chemokine receptor type 4 (P70658)

1      MEPISVSIYT SDNYSEEVGS GDYDSNKEPC FRDENVHFNR IFLPTIYFII FLTGIVGNGL
61     VILVMGYQKK LRSMTDKYRL HLSVADLLFV ITLPFWAVDA MADWYFGKFL CKAVHIIYTV
121    NLYSSVLILA FISLDRYLAI VHATNSQRPR KLLAEKAVYV GVWIPALLLT IPDFIFADVS
181    QGDISQGDDR YICDRLYPDS LWMVVFQFQH IMVGLILPGI VILSCYCIII SKLSHSKGHQ
241    KRKALKTTVI LILAFFACWL PYYVGISIDS FILLGVIKQG CDFESIVHKW ISITEALAFF
301    HCCLNPILYA FLGAKFKSSA QHALNSMSRG SSLKILSKGK RGGHSSVSTE SESSSFHSS
```

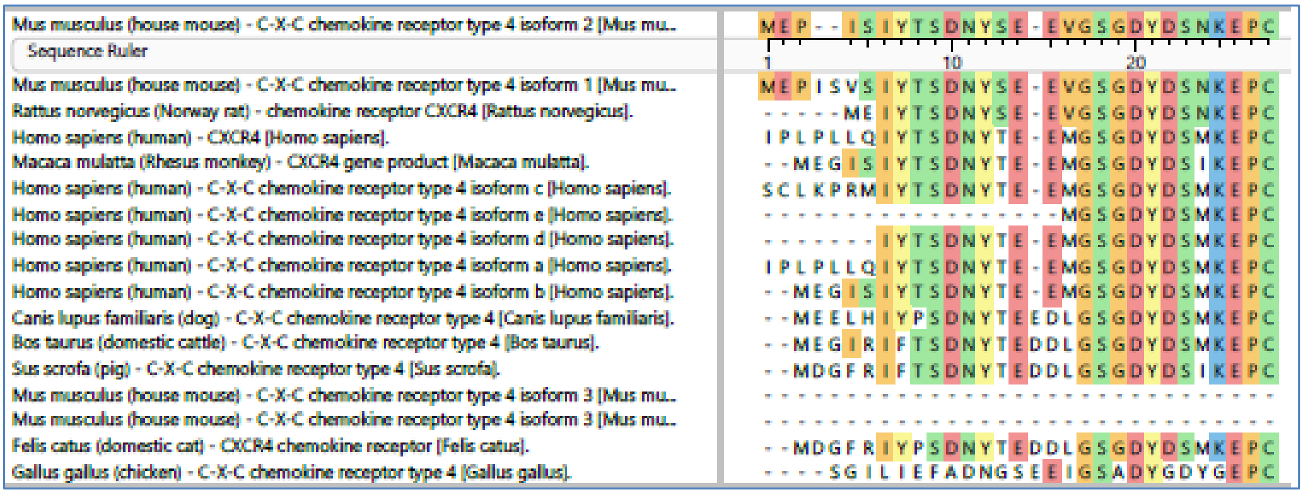
CD184 epitope in structures

CD184 or CXCR4 is a transmembrane receptor and the antibody recognizes the extracellular N-terminus. Therefore, no structures of the full protein have been published. The picture below shows the otherwise unstructured N-terminal sequence (yellow) in complex with a ligand. **SIYTSDnYsEE** side chains are displayed as spheres, the ligand as green cartoon. This structure (PDB 2k04) shows the N-terminus of the human protein. It can be seen how the amino acids forming the epitope could be accessed by an antibody. This structural view explains also, why the epitope is not a continuous sequence. The specificity of this antibody results not from a defined structure, but from the number of conserved residues recognized in a (without ligand) disordered structure.



Species specificity

The antibody originating from rat is well described as specific for the mouse protein. Nevertheless, the N-terminus of the CXCR4 is well-conserved among different species.



Mus musculus (house mouse) - C-X-C chemokine receptor type 4 isoform 2 [Mus mu...	MEP--ISIYTSDNYSE-EVGSGDYDSNKEPC
Sequence Ruler	1 10 20
Mus musculus (house mouse) - C-X-C chemokine receptor type 4 isoform 1 [Mus mu...	MEPISVSIYTSDNYSE-EVGSGDYDSNKEPC
Rattus norvegicus (Norway rat) - chemokine receptor CXCR4 [Rattus norvegicus].	-----MEIYTSDNYSE-EVGSGDYDSNKEPC
Homo sapiens (human) - CXCR4 [Homo sapiens].	IPLPLLQIYTSDNYTE-EMGSGDYDSMKEPC
Macaca mulatta (Rhesus monkey) - CXCR4 gene product [Macaca mulatta].	--MEGISIYTSDNYTE-EMGSGDYDSIKEPC
Homo sapiens (human) - C-X-C chemokine receptor type 4 isoform c [Homo sapiens].	SCLKPRMIYTSDNYTE-EMGSGDYDSMKEPC
Homo sapiens (human) - C-X-C chemokine receptor type 4 isoform e [Homo sapiens].	------------------MGSGDYDSMKEPC
Homo sapiens (human) - C-X-C chemokine receptor type 4 isoform d [Homo sapiens].	-------IYTSDNYTE-EMGSGDYDSMKEPC
Homo sapiens (human) - C-X-C chemokine receptor type 4 isoform a [Homo sapiens].	IPLPLLQIYTSDNYTE-EMGSGDYDSMKEPC
Homo sapiens (human) - C-X-C chemokine receptor type 4 isoform b [Homo sapiens].	--MEGISIYTSDNYTE-EMGSGDYDSMKEPC
Canis lupus familiaris (dog) - C-X-C chemokine receptor type 4 [Canis lupus familiaris].	--MEELHIYPSDNYTEEDLGSGDYDSMKEPC
Bos taurus (domestic cattle) - C-X-C chemokine receptor type 4 [Bos taurus].	--MEGIRIFTSDNYTEDDLGSGDYDSMKEPC
Sus scrofa (pig) - C-X-C chemokine receptor type 4 [Sus scrofa].	--MDGFRIFTSDNYTEDDLGSGDYDSIKEPC
Mus musculus (house mouse) - C-X-C chemokine receptor type 4 isoform 3 [Mus mu...	-------------------------------
Mus musculus (house mouse) - C-X-C chemokine receptor type 4 isoform 3 [Mus mu...	-------------------------------
Felis catus (domestic cat) - CXCR4 chemokine receptor [Felis catus].	--MDGFRIYPSDNYTEDDLGSGDYDSMKEPC
Gallus gallus (chicken) - C-X-C chemokine receptor type 4 [Gallus gallus].	----SGILIEFADNGSEEIGSADYGDYGEPC

According to this alignment it is most likely, that N- and C-terminus of the epitope recognized by the antibody are important for this specificity, while the central region is shared amongst the different receptor proteins.

Contact:

epitopic GmbH – Deutscher Platz 5e - 04103 Leipzig - Germany

E-Mail: info@epitopic.com

Phone +49 341 253 55 160

www.epitopic.com