

Epitope Mapping report

Date: 05.03.2015

Target: purified Mouse Anti-Human mAB CD227 (BD Biosciences, Clone HMPV)

Antigen: CD227 / Mucin-1

Panning conditions: Immobilized on M270 carboxy beads

NGS library statistics:

Dataset:	CD227-ENTE12016-1pr (1st Panning)
Number of sequences:	607,582
Number of valid sequences:	377,990
Pattern/Library:	ENTE-1
3-mer motifs:	7,254
4-mer motifs:	131,956

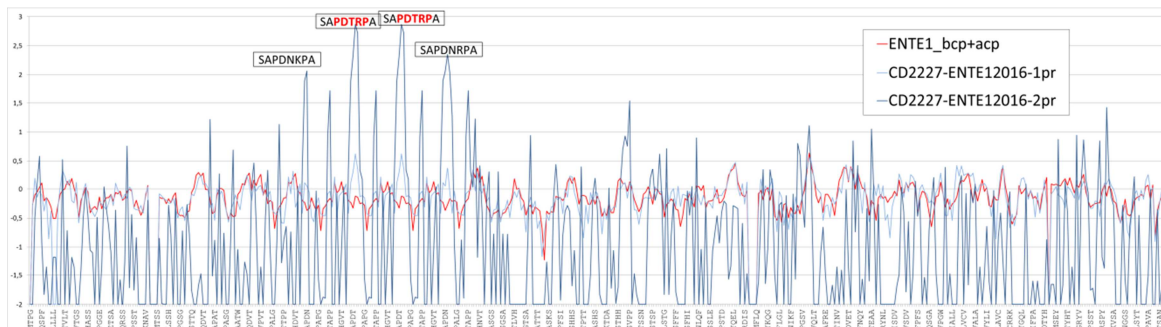
Number of sequences is the output from NGS

Valid sequences are those finally fitting all QC standards and used for the analysis

Dataset:	CD227-ENTE12016-2pr (2nd Panning)
Number of sequences:	649,316
Number of valid sequences:	410,803
Pattern/Library:	ENTE-1
3-mer motifs:	7,215
4-mer motifs:	77,530

The number of *motifs* in the first round should be close to theory, i.e. ca 7,200 resp. 132,000 for medium size data sets of up to 500,000 sequences

Motif statistics:



Statistic of 4mer motifs in the data sets against a short version of MUC1, which has the identical repeats being removed. The reference data set contains almost 2 Mio Sequences from the naïve library.

X-axis: All possible 4mer fragments of the antigen starting from 1-4, 2-5 ...to... 489-492

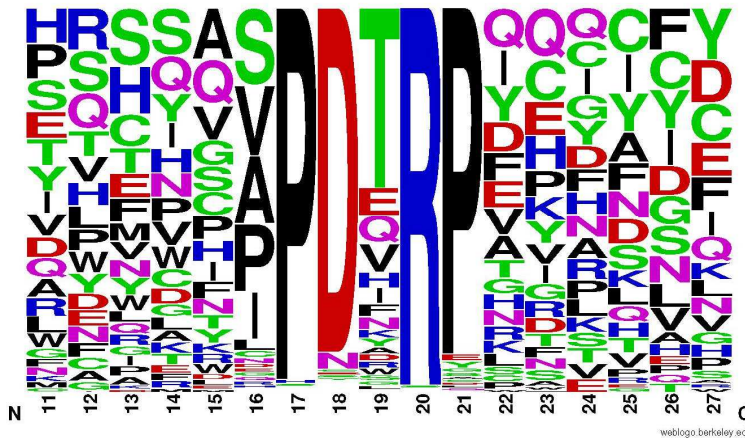
Y-axis: log of enrichment vs. calculated probability in the data set, e.g. 3 is 10^3 resp. 1000-fold enrichment. Control ENTE1 naïve starting library has max 0.5 deviation from 0. After the first panning round the library is depleted of motifs, many approach a cut off of <-2 , which should not be taken 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:

PD(T)RP

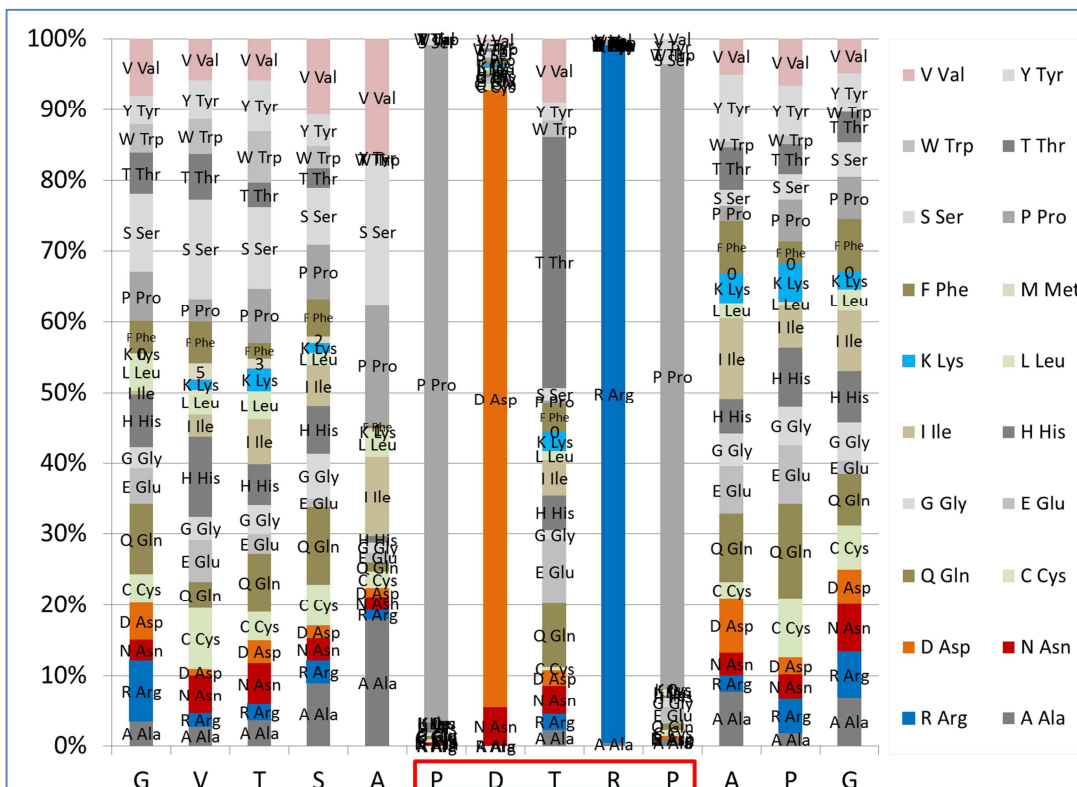
Presentation as web logo (all 223 sequences of the first selection round containing at least 4 amino acids out of PDTRP)



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

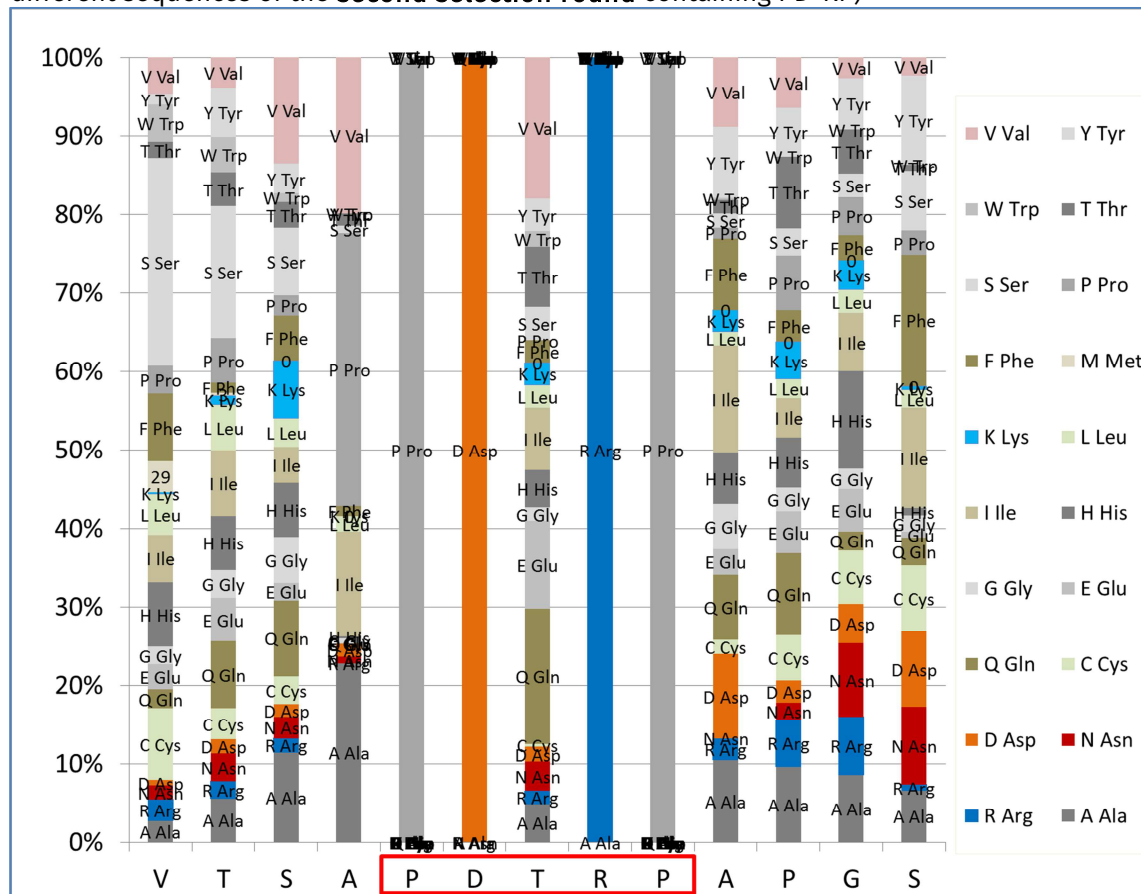
The web-logo is available online:
<http://weblogo.berkeley.edu/>

Presentation as EPITOPIC fingerprint (all 223 sequences of the **first selection round** containing at least 4 amino acids out of PDTRP)

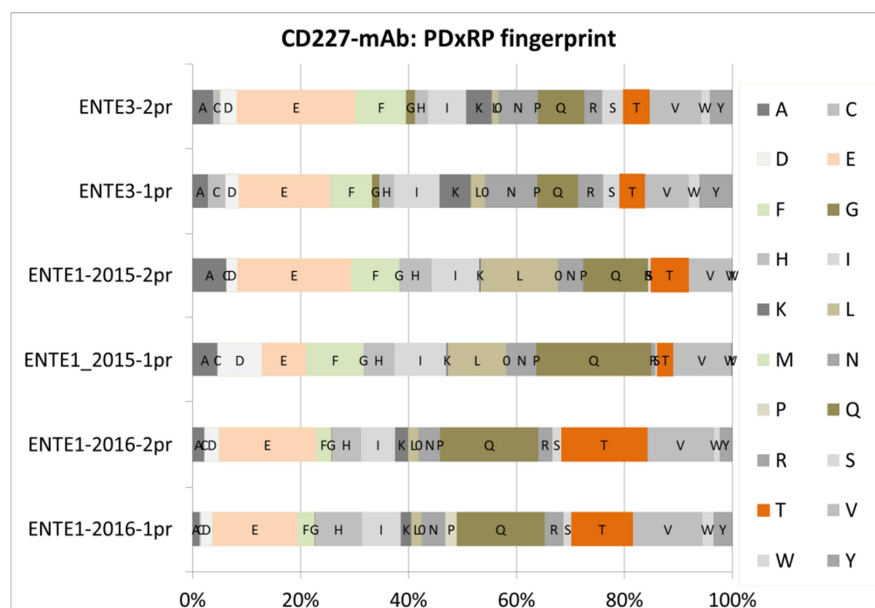


Fingerprint is the relative frequency of amino acids observed close to a search sequence. Fingerprints may vary with the search sequence. Usually repeated runs with the ENTE-1 library result in similar fingerprints. The motif fingerprint of an antibody is therefore as reliable as a human fingerprint.

Presentation as EPITOPIC fingerprint (from 63,248 sequences containing 1,014 different sequences of the **second selection round** containing PD*RP)



Statistics for the x in the PDxRP Motif from several control experiments including test runs from the new library ENTE3. This reveals a preference for glutamic acid over the threonine in the centre of the naïve sequence epitope.



Epitope in Sequence (the identical sequence repeats have been reduced to 2)

MTPGTQSPFFLLLLLTVLTVVTGSGHASSTPGGEKETSATQRSSVPSSTEKNAVSMTSSVLSSHSPGSGSSTTQGQDVT
LAPATEPASGSAATWGQDVTSVPVTRPALGSTTPPAHDVTSAPPDNKPAPGSTAPPAHGVTSAPPDTRPAPGSTAPPAHGV
TSAPPDTRPAPGSTAPPAHGVTSAPPDNRPALGSTAPPVHNVTASGSASGSASTLVHNGTSARATTPASKSTPFSIPSH
HSDTPPTTLASHSTKTDASSTHHSSVPPLTSSNHSTSPQLSTGVSFFFFLSFHISNLQFNSSLEDPSTDYYQELQRDISEM
FLQIYKQGGFLGLSNIKFRPGSVVVQLTLAFREGTINVHDVETQFNQYKTEAASRYNLTISDVSVDVPPFSAQSGAG
VPGWGIALLLVLCVLVALAIVYLIALAVCQCRRKNYGQLDIFPARDTYHPMSEYPTYHTHGRYVPPSSTDSPYEKVS
A
GNGGSSLSYTNPAVAATSANL

The sequence labelled in red is the most likely epitope of the antibody. The normal MUC-1 protein contains a gene multiplication of this motif. Other similar motifs are unlikely to be the sequence the antibody was raised against.

Sequence Examples: Alignment of all 224 sequences from the first selection round with at least four out of the amino acids of the PDTRP-motif.

MUC-1	AHGVTSAPDTRPAPGST	90	CTDY YGP PDQ RPI	180	VDASV APDKRPAEFS
1	--AGACV PDERF GITQA	91	KDRCQIP PDTRL PFI	181	VDSSIVP PDARPD AHN
2	AGTCHIP PDTRYF PRD	92	KIRSQPL PLTRP LEID	182	VDTSRI PDRRP ESRRQ
3	--AIGSI PDQRP IRPSG	93	KIRSQPS --TRPKGEDN	183	ESCVEYD PDERP QEY
4	--AKWS PDV RPTIFAS	94	--KNVS PDH RPYGIQS	184	SFDRAG APDGRPD
5	APSSFV PDIRF GVSVC	95	--KQGS PDFRPVQAHP	185	CPVTWPV PDFRPD
6	RASEHGV PDWRPYQA	96	--LDN SL PDTRYQNVSP	186	CTAKYQV PDARPF
7	ARQSGH PDTA PNVRRN	97	LFQCQQA PDTRGKNDY	187	SQEAHHI PDFRPF
8	CKTWLF ADTRCF	98	CYNHVGI PDTRPY	188	--VKRS PDIRP YEKHV
9	--DETCF PDRRPYFEHG	99	--LNFCD PTDVPHCSQE	189	--VLRCA PDVRPEGHKN
10	CVDPNSA PNTTRPF	100	QGSWWNA PDRRPYEA	190	YAYIYAQ PDTRD
11	GQCFEWR PDYRPEQC	101	--LQVS PDQRPQVVV	191	QSHDTP PDWRP GNF
12	HACPPV PDNRP THHC	102	SWLESAI PDERPN	192	QSSNWV APDGRPSHC
13	HCHQA PDTRP EKYC	103	LTMSKAPDTRPVQKIY	193	VQWCRVP PDVRP RKDDC
14	--DIACDTRPERGSD	104	VASTDHA PDWRPVQC	194	SPRHHFV PDVRPD
15	--DPM SAPNTRPKEDTA	105	--LWHS PDSRPHQHVV	195	VMSST SAPDWRPPND
16	SYLIIDP PDERPA	106	YNSFIFV PDQRPRI	196	VVCEV PDERP VVNC
17	SEQDRGV PDFRPD	107	--NAHS PDQRPICILQ	197	--VVVCP ETRPEQYH
18	DRVSLQ PDIRP PNA	108	DWCEIKV PDERPFI	198	WASGNTN PDVRPHY
19	DVFSGW PDKRPQQF	109	--NET SAPDERPPVGYT	199	QSWQVVP PDQRPYTA
20	CTWHNAV PDTRPF	110	--NFMSPDNRRPRGII	200	--VWS SI PDNRP IDILP
21	EACNP ADTRLITPF	111	CERNHYGPH TRPI	201	VWTCHP PDYRPSNNA
22	EDGSYPD PDTRAEVYD	112	CHTSHIP PDQRPF	202	CSIFWPL PDERPI
23	EDRSPGI PDIRP PNA	113	HMCSSHV PDERPSKC	203	--WDHS PDTRGQGIS
24	EWSKIP PDTRHKTGI	114	--NHSSP NTRPEQSQG	204	WLGSPYP PDQRP IQHD
25	--EFMSEH DTRPEKCDY	115	--NIQSP DTRQPYGSI	205	WRFCFRP PYTRPDQA
26	EHASVP PDSRP QCHYC	116	--NKNS PDFRPVWTFD	206	YAQSWQV PDERPDYPI
27	--EHMS PNTTRP CTVN	117	SPRFYAP PDTRPN	207	YDWCWNI PNTRPAHVY
28	EINSAQP PDVRP PCTY	118	--NPPSPDWRPSTII	208	--YEFSP DERPA CPTC
29	CNLFPPP PDV RPF	119	CVSFLVP PDARPI	209	--YFQSP DQRP HCCYS
30	--ENMSYN PYTRPLDINY	120	NQCCSEQ PDGRPEKRA	210	--YFQSP DQRP HCCQA
31	NVSP LGADKRP IYD	121	NYLSFA PDV RPLITPY	211	YFQSWDP PDDRP RVPF
32	--EQHC PNTTRP PVT	122	YWCEN SP PDIRPDQI	212	YNNCF ADTREYQVQC
33	--EQQS PDTRP PONG	123	PAWSVV PDERPVHNSY	213	YNTCLNDTRPCKQST
34	VSRFRV PDCKRPQQA	124	SPLWLL APNTRPI	214	--YNVSP DNRPTIFLF
35	CEFPYH PDNRPC	125	PDQSTHPPD FRPAFAY	215	--YNVSP DNRPTIQLH
36	--EYAS PDDRPTIKVQ	126	STHIYYP DLRPA	216	CVWWGQI PDDRPF
37	EYMC P DHRPGHSAI	127	--PHQSVP DNRPRQPHF	217	MCTVA VPIITRPTA
38	EYPCWFP DTRSPVDP	128	CLVALNR PDTRSF	218	SWELPLA PDTR EY
39	CHYVQH I PDQRP I	129	--PPQC PFTRPEYAHH	219	--YRQCP PDTRPQHGA
40	SWVHQAI PDIRPA	130	QMCWLRG PDIRPVKD	220	YRWSYPR PSTRPHRIC
41	CGHNRPN PNTTRPC	131	QSEEHPP DRRPTIA	221	SKLHTAP PDYRPD
42	SITVHHH ADNRPA	132	CAHTRCP PDVRPA	222	YTPSYGV PDVRPCGNA
43	--FLVS PDTRPAIFYS	133	QCQWQP PETRPIN	223	YTPSYGV PDVRPKYTI
44	QWSPIDI PDERP NPGF	134	SVVLPFGP PTRPD	224	YWHSSV PDIRP QFYDN
45	CHYQPVN PDTRQY	135	SYVQQIY PDTRSD		
46	HESEAI PDVRPA	136	EQSRGVAP DIRPQRI		
47	FESVKVP PDTRY	137	--QFPCLP DERPITATF		
48	GERCYV ADHRRPKHNF	138	--QFPCLP DKRPITATF		
49	GERCYV ADHRRPKQNF	139	--QHVSVP DARPWSHDI		
50	--GNAS PDQRP HQHRN	140	CLGYYSIPNTRPY		
51	SEPIWSV PDQRPY	141	--QVVSAPD TREYVPC		
52	GPCHYV ADVRP GSNF	142	--QYACQD TRPLWSPV		
53	SREQVAE PDTRGY	143	CSQNGI IPDFRPA		
54	GQMSHA PDDRPKTQVF	144	RGTSVPH DTRPNQLYI		
55	QFW --DTRPRLPFF	145	RHGSLAP DTRCHRRYC		
56	--GVWS PDQRP YQDKF	146	SYEVP AQPVTRPA		
57	SKVG IYL PDERN	147	--RRMCYP DHRPNYRD		
58	--HDQS PDERP IGFA	148	--RTQS IPDHRPGYLAG		
59	--HDQS PDERP IGRID	149	CPEYCF A		
60	--HEAS PDHRRPVCSLN	150	SEHSQYES DTRPVQSY		
61	STTLQVI PDV RPF	151	CSQQTVVP DIRPF		
62	CNEQGAS PDVRPD	152	--SGQS PDHRRPTILTV		
63	SDHEHPS PNTRPY	153	--SGQS PDHRRPTIRGD		
64	--HHAS PNTRP YLEF	154	SGTCWIP PDDRPKSRN		
65	HLSSSTV PDTRPVCLSC	155	SIQSINI PDTRPDCCI		
66	--HLVS PDRRPIQTQG	156	SPSSA APDLRPQHQC		
67	HPFSPGI PDVRPERFC	157	SRPCQA PDVRPHCHIC		
68	--HPHS PDERP SQAHN	158	SSMCAV PDTRPLEDCI		
69	HPVCHGVH DTRPTHC	159	--SVPS TRPNFYIY		
70	HRFSQI PDYRPTPLD	160	SRQHIEAPDQRPF		
71	--HVAS PDSRPEEGTY	161	--TAQSPDQRPICGAE		
72	--HVQS PDLRPTFLE	162	--TAQSPDQRPICILQ		
73	ESLAIAP PDTRY	163	--TDAS PDTRPRCEKL		
74	--HWQS PDIRPQENYT	164	CEVNNQA PDTRREF		
75	--HYS CAPDTRAGKGEV	165	TDVSWSP NTRPQEHF		
76	--HYS CAPDTRAGSHNV	166	THFSPP PDHRRPECLY		
77	--HYVS APDTRYEYQFN	167	--THHCVP DKRPHQHYF		
78	SYLFPLV PDVRPD	168	--THHCVP DNRPHQHYF		
79	IDHCWA PKTRPDVQDD	169	CPRGGIV PDQRPY		
80	CWHESTI PDVRPY	170	IPSIHHP DERRPDN		
81	SDHVP ANTRPD	171	IQCTGDPDPRPFN		
82	VSYPYIV PCTRPID	172	--TLPC TRPYEKPV		
83	--IPMS DTRPVHCQL	173	CSPHII IPDERPA		
84	IPPCIIV PDVRPTTYF	174	TSVCNTP DSRPYQRD		
85	IPQSVII PDFRPVHSI	175	--TVPS PDQRPHPVCC		
86	IQFSPYNG DTRPSCYY	176	--TWRS PDYRPEGHRY		
87	--IQVCTV PDVRPGQYNF	177	--TWWS PDKRPAQPKH		
88	ITFSVKP PDIRPVERF	178	VAPCKDV PDARPAHV		
89	CDPINF PDFRPI	179	--VAVS PDHRRPGF		

The motif search can reveal very large numbers of similar sequences. The data presented here is only an excerpt, based on a search for a restricted sequence motif. Broader searches can reveal thousands of such sequences in the first selection round and hundreds of thousands from second selection rounds on monoclonal antibodies.