



Supplemental Figure 1. Comparison of beta chain CDR3 of TCR either crystallized as part of a pMHC complex or unliganded. CDR3 amino acid position 4 is circled. The cysteine identifying the beginning of the CDR3 and the phenylalanine from the conserved FG motif in the J region, signifying the end of the CDR3 are identified in panel A. all three CDR3 structures are oriented in a similar manner with the V-derived portion of the CDR3 in the foreground. The peptide from the complex is also shown to provide a sense of perspective. The right angle turn at position 4 is evident in all three structures. A. The RL42 TCR recognizing EBNA3 in the context of HLA-B8 (green mainchain), or unliganded (pink mainchain). The complex structure is from PDB file 3SJV and the unliganded structure from file 3SKN (Gras et al. *J Immunol.* 188: 311,2012). B. The DMF5 TCR recognizing the MART peptide in the context of HLA-A2 (green mainchain), or unliganded (pink mainchain). The complex structure is from PDB file 3QDG and the unliganded structure from file 3QEU (Borbulevych et al. *J Immunol.* 187:2453, 2011). C. L13 TCR recognizing EBNA3 in the context of HLA-B8 (green mainchain), or unliganded (blue mainchain). Complex structure from PDB file 1M15 (Kjer-Nielsen et al. *Immunity* 18:53, 2003), unliganded structure from file 1KGC (Kjer-Nielsen et al. *Structure* 10: 1521, 2002).

A

	CDR3 amino acid position															percent
	2	3	4	5	6	7	8	9	10	11	12	13	14	15		
G	0.07	0.13	0.3	4.05	3.91	5.57	4.21	2.63	1.07	0.49	0.18	0.05	0.02	0.01	22.725	
R	0.01	0.52	1.53	1.65	2.33	1.73	1.25	0.82	0.5	0.27	0.1	0.05	0.03	0	10.793	
S	0.12	0.68	0.39	1.02	1.53	1.52	1.27	0.84	0.55	0.26	0.11	0.07	0.02	0.01	8.375	
A	0.01	0.04	0.2	1.4	1.37	1.76	1.45	0.87	0.45	0.19	0.08	0.03	0.01	0	7.872	
P	0	0.02	2.39	0.69	1.34	0.63	0.59	0.45	0.28	0.14	0.08	0.02	0.01	0.01	6.648	
T	0.06	0.27	0.2	1.21	1.74	1.21	0.84	0.56	0.25	0.12	0.1	0.03	0.02	0.02	6.621	
L	0	0.01	1.29	0.81	1.64	0.9	0.75	0.54	0.33	0.12	0.07	0.04	0.01	0.01	6.522	
D	0	0	0.13	3.27	0.74	0.7	0.55	0.42	0.28	0.13	0.06	0.02	0.02	0.01	6.323	
Q	0.01	0.09	1.53	0.61	0.99	0.75	0.47	0.19	0.14	0.05	0.03	0.01	0	0	4.864	
V	0.01	0.02	0.1	1.39	0.75	0.65	0.68	0.46	0.23	0.12	0.05	0.02	0.01	0	4.473	
E	0	0.02	0.12	1.31	0.4	0.38	0.43	0.28	0.14	0.11	0.05	0.02	0	0	3.284	
H	0	0.01	0.93	0.14	0.2	0.14	0.16	0.23	0.16	0.08	0.04	0.02	0	0	2.114	
N	0.03	0.09	0.08	0.32	0.25	0.26	0.28	0.26	0.13	0.07	0.03	0.02	0.01	0	1.832	
I	0.01	0.02	0.07	0.22	0.32	0.2	0.26	0.23	0.14	0.07	0.03	0.01	0	0	1.578	
F	0	0	0.13	0.19	0.31	0.18	0.23	0.21	0.11	0.06	0.02	0.02	0	0	1.477	
K	0.01	0.05	0.17	0.25	0.19	0.18	0.18	0.16	0.1	0.09	0.03	0.02	0	0	1.432	
Y	0	0.01	0.1	0.19	0.17	0.13	0.24	0.18	0.13	0.08	0.04	0.02	0.01	0.01	1.303	
W	0	0	0.04	0.17	0.33	0.19	0.15	0.1	0.07	0.03	0.01	0	0	0	1.102	
M	0	0.01	0.03	0.12	0.12	0.08	0.09	0.06	0.05	0.02	0.01	0.01	0	0	0.589	
C	0	0	0	0.01	0.02	0	0	0.01	0.01	0	0.01	0.01	0	0	0.073	
	total observations														50433	

B

	CDR3 amino acid position															percent
	2	3	4	5	6	7	8	9	10	11	12	13	14	15		
G	0.33	0.58	1.33	4.37	6.21	4.58	3.56	2.13	0.9	0.35	0.13	0.06	0.01	0.01	24.548	
R	0.1	2.19	1.17	1.79	1.79	1.65	1.2	0.81	0.49	0.21	0.08	0.03	0.01	0	11.53	
P	0	0.13	3.49	2.69	1.08	0.93	1.03	0.5	0.28	0.18	0.1	0.03	0.01	0	10.463	
S	0.23	2.68	1.68	1.29	0.81	0.91	0.7	0.56	0.32	0.22	0.07	0.07	0.05	0.01	9.581	
T	0.5	1.62	0.86	0.97	1.17	1.12	0.57	0.38	0.21	0.11	0.1	0.05	0.02	0.01	7.7	
A	0.06	0.15	0.54	1.09	1.23	1.41	1.12	0.7	0.37	0.15	0.08	0.03	0.01	0	6.932	
L	0.03	0.19	2.09	0.95	0.8	0.73	0.65	0.31	0.21	0.1	0.06	0.03	0.02	0.01	6.172	
Q	0.02	0.23	0.44	0.72	0.72	0.61	0.26	0.16	0.1	0.07	0.02	0.01	0.01	0	3.362	
V	0.02	0.07	0.25	0.53	0.53	0.63	0.56	0.43	0.15	0.09	0.05	0.03	0	0	3.362	
D	0	0.03	0.25	0.46	0.6	0.49	0.35	0.28	0.19	0.07	0.06	0.01	0.01	0	2.812	
F	0	0.01	1.55	0.27	0.16	0.18	0.15	0.15	0.1	0.03	0.01	0	0.01	0	2.622	
E	0.01	0.06	0.22	0.21	0.27	0.37	0.21	0.18	0.15	0.08	0.05	0.01	0	0	1.819	
W	0	0.04	0.31	0.52	0.21	0.2	0.2	0.11	0.07	0.03	0.01	0	0	0	1.704	
H	0.01	0.14	0.22	0.23	0.15	0.2	0.2	0.16	0.1	0.03	0.01	0.01	0	0	1.469	
N	0.08	0.28	0.12	0.19	0.14	0.14	0.17	0.18	0.09	0.05	0.02	0	0	0	1.46	
Y	0	0.05	0.19	0.21	0.1	0.19	0.22	0.22	0.1	0.06	0.04	0.01	0.01	0	1.403	
I	0.05	0.24	0.17	0.21	0.14	0.14	0.16	0.1	0.09	0.04	0.02	0	0	0	1.347	
K	0.02	0.17	0.16	0.1	0.08	0.12	0.13	0.14	0.07	0.05	0.02	0.01	0	0	1.078	
M	0	0.07	0.04	0.09	0.07	0.05	0.07	0.06	0.04	0.01	0.01	0	0	0	0.517	
C	0.04	0.02	0.01	0.01	0	0	0	0.01	0	0.01	0	0.01	0	0	0.12	
	total observations														46842	

Supplemental Figure 2. Amino usage for the two CD4 clonotype data sets. A. BV3. B. BV28. Relative frequency is shown as percent of the use of that amino acid at the CDR3 position defined above each column. The total number of amino acid observations used as the denominator is shown at the bottom right. The last column summarizes the amino acid use for the entire NDN dataset. Amino acids that are V encoded at CDR3 positions 1 to 5 are boxed. The frequent occurrence of the same a.a. in the NDN is due to synonymous substitution.