Enhancing and inhibitory motifs have coevolved to regulate CD4 activity.

Mark S. Lee[1], Peter J. Tuohy[1], Caleb Kim[1], Katrina Lichauco[1], Heather L. Parrish[1], Koenraad Van Doorslaer[1, 2, 3, 4, 5*] and Michael S. Kuhns[1, 3, 4, 5, 6*]

[1] Department of Immunobiology, The University of Arizona College of Medicine, Tucson, AZ 85724, USA.

[2] School of Animal and Comparative Biomedical Sciences, The University of Arizona, Tucson, AZ 85724, USA.

[3] Cancer Biology Graduate Interdisciplinary Program and Genetics Graduate Interdisciplinary Program, The University of Arizona, Tucson, AZ 85724, USA.

[4] The BIO-5 Institute, The University of Arizona, Tucson, AZ 85724, USA.

[5] The University of Arizona Cancer Center, Tucson, AZ, USA

[6] The Arizona Center on Aging, The University of Arizona College of Medicine, Tucson, AZ, USA

[*]Correspondence: vandoorslaer@arizona.edu, mkuhns@email.arizona.edu

Running title: Coevolved motifs under purifying selection regulate CD4 activity

SUMMARY

$CD4^+$ T cells use T cell receptor (TCR)-CD3 complexes, and CD4, to respond to peptide antigens within MHCII molecules (pMHCII). We report here that, through ~435 million years of evolution in jawed vertebrates, purifying selection has shaped motifs in the extracellular, transmembrane, and intracellular domains of eutherian CD4 that both enhance pMHCII responses and are coevolving with residues in an intracellular motif that inhibits pMHCII responses. Importantly, while CD4 interactions with the Src kinase, Lck, are classically viewed as the key determinant of CD4's contribution to pMHCII responses, we found that without the inhibitory motif CD4-Lck interactions are not necessary for robust responses to pMHCII. In summary, motifs that mediate events on the outside and inside of $CD4^+$ T cells coevolved to finetune the relay of pMHCII-specific information across the membrane. These results have implications for the evolution and function of complex transmembrane receptors and for biomimetic engineering.

46 **INTRODUCTION**

47 The immunological 'Big Bang' that gave rise to RAG-based antigen receptor gene
48 rearrangement in jawed vertebrates produced an adaptive immune system in which
49 each naïve B and T cell expresses a clonotypic B or T cell receptor (BCR or TCR)
50 encoded by gene segments rearranged by Rag-1 and Rag-2 (Bernstein et al., 1996).
51 Most T cells use their TCRs together with either CD4 or CD8 coreceptors to recognize
52 peptide antigens presented by MHCII or MHCI (pMHCII or pMHCI) molecules.
53 Interestingly, *Callorhinchus milii*, a cartilaginous fish with an immune system that is
54 thought to resemble the primordial adaptive immune system, encodes BCRs, TCRs,
55 MHCI, MHCII, and CD8 within its genome – but an orthologous gene encoding CD4
56 appears to be absent, as are genes for proteins associated with $CD4^+$ T cell helper (Th)
57 or regulatory (Treg) functions (e.g. *FoxP3* and *Rorc*) (Venkatesh et al., 2014). These
58 findings were taken as evidence that CD4 evolved to help direct more elaborate T cell
59 functions than could be achieved by TCR engagement of pMHCII alone.

60 While this idea is intriguing, most models attribute the quantity and quality of
61 pMHCII-specific signaling to the biophysical properties of TCR-pMHCII interactions;
62 they then assign CD4 a supporting role whereby it interacts with the Src kinase, Lck, via
63 an intracellular CQC clasp motif to recruit Lck to phosphorylate the immunoreceptor
64 tyrosine-based activation motifs (ITAMs) of the TCR-associated CD3 signaling modules
65 when both the TCR and CD4 coincidently engage pMHCII (Glaichenhaus et al., 1991;
66 Guy and Vignali, 2009; Kuhns and Davis, 2012; Malissen and Bongrand, 2015; Turner
67 et al., 1990; Yin et al., 2012). It is challenging to envision how CD4 might direct
68 elaborate pMHCII responses based on these models.

115 >1 and those under purifying selection have a dN:dS ratio <1 (**Figure S1C and S1D**).
116 Of the 17 codons under diversifying selection, 16 (94.1%) are distributed across the
117 CD4 ECD while only one is found in the TMD. Of the 126 residues under purifying
118 selection, 98 are distributed across the CD4 ECD (24.8% of all codons in the ECD). In
119 contrast, 45.5% of TMD codons (10 of 22) and 45% (18 of 40) within the ICD are under
120 purifying selection. In sum, the ECD contains residues under selective pressure due
121 either to their structural or functional importance, while the TMD and ICD of mammalian
122 CD4 are particularly constrained due to their importance. As the ICD of human CD4 is
123 largely unstructured, the residues in the ICD are likely to be functionally important (Kim
124 et al., 2003).

125 We proceeded to focus on an ECD region that includes the D3 domain solvent-
126 exposed nonpolar patch that enhances CD4 function, the TMD, and the ICD to more
127 closely interrogate their evolution (Glassman et al., 2018). Accordingly, we generated a
128 maximum likelihood phylogenetic tree that included the predicted most recent common
129 ancestor (MRCA) sequences at each node to evaluate the evolutionary events that
130 dictate structure and function at our regions of interest (**Figure 1A and Figure**
131 **S1E**)(Hochberg and Thornton, 2017). We also evaluated residue occurrence in
132 eutherians by logo plot analysis to better consider positional variability of residues within
133 this clade. Additionally, we related evolutionary changes on a residue-by-residue basis
134 to our FEL results for all extant CD4 molecules, as well as the mammalian only dataset,
135 to gain a more complete picture of the pressures shaping CD4 molecules (**Figure 1B**
136 **and 1C**). Particular attention was given to mouse and human CD4 due to their
137 experimental and human health relevance.

138 First, we focused on the evolutionary history of the PLXF motif (mouse 228-231)
139 and P281 residue in the D3 domain of the CD4 ECD to ask if our analyses would
140 highlight residues that are adjacent in 3D space, form a solvent-exposed nonpolar
141 patch, facilitate the formation and stability of TCR-CD3-pMHCII-CD4 assemblies, and
142 increase responses to weak and agonist pMHCII (Glassman et al., 2018; Wu et al.,
143 1997). The MRCA of all amniotes contains a PLXF motif (mouse 228-231) in the D3
144 domain and is fixed in mammals (**Figure 1A,** node 1). P281 is not found in the amniote
145 MRCA but is fixed in all mammals, suggesting that it arose later. Importantly, P228,
146 L229, F231, and P281 all have very small dN:dS values that are primarily driven by low
147 dN rates, suggesting that they are of structural or functional importance. Indeed,
148 structural analysis indicates that L229 is buried in the hydrophobic core of the D3
149 domain as is L282 adjacent to P281, implying a structural function, while the solvent
150 exposed P228, F231, and P281 are known to be important for CD4 function, providing a
151 possible explanation for why these residues are under purifying selection (Glassman et
152 al., 2018; Wu et al., 1997).
153 We next turned our attention to residues within the TMD and ICD by focusing on
154 motifs that: 1) displayed a stepwise refinement from ancestral sequences towards
155 extant sequences; 2) contain highly conserved residues with evidence of purifying
156 selection. For example, the MRCA of all amniotes (node 1) contains a GG patch (G402
157 and G403) in its TMD that is lost in non-avian reptiles but present in extant birds (**Figure**
158 **1A**). In the eutherian MRCA (node 4) the GG patch expanded into a highly conserved
159 GGXXG motif that is present in the majority of eutherian CD4 proteins (see logo plot)
160 including mouse and human. The selection analysis also identified G402 and G403

161 using both the full and mammal-only dataset, while G406 was only identified in the full
162 dataset (**Figure 1B and 1C**). As we have ruled out a role for the GGXXG motif in
163 mediating mouse CD4 dimerization, it is likely to mediate heterotypic protein-protein or
164 protein-cholesterol interactions (Parrish et al., 2015; Teese and Langosch, 2015).
165 Using our approach we also identified a CV+C motif ("+" represents a basic
166 residue) that fixated in eutherians (**Figure 1A-1C**). Palmitoylation of these cysteines is
167 reported to influence membrane raft localization, although this is controversial as
168 association with Lck may also localize CD4 to membrane rafts (Crise and Rose, 1992;
169 Fragoso et al., 2003; Ladygina et al., 2011). Our analysis suggests that the combination
170 of the CV+C and the GGXXG/S motifs are unique to extant eutherians, co-arose during
171 evolution, and may work together to regulate CD4 membrane localization.
172 A poly-basic RHRRR motif in the juxtamembrane region of the ICD was also
173 previously reported to impact human CD4 localization to membrane rafts (Popik and
174 Alce, 2004). Our comparative analyses suggest a core HXXR motif (mouse 423-426)
175 with H423 and R426 under purifying selection (**Figure 1A-1C**).
176 Further downstream, NMR has shown that the ICD of human CD4 contains a
177 helix-turn structure, the sequence of which is highly conserved in all mammalian CD4
178 molecules (grey shaded region, 429-442, **Figure 1A**)(Kim et al., 2003; Willbold and
179 Rosch, 1996). A conserved IKRLL motif, resembling known I/LXXL protein-interaction
180 motifs, is embedded within the helix. Its origins are evident in the amniote MRCA (node
181 1) via the presence of a dileucine repeat. Reptiles and birds diverged away from this
182 motif, while the MRCA of mammals (node 3) evolved a conserved IXRLL. This region
183 includes a number of residues under purifying selection within the full dataset analysis

184 (R430, 434-438 that make up the IKRLL motif, S439, and 440-442 that form the turn) or
185 within the more limited mammalian dataset (436-440 and 442) (**Figure 1B and 1C)**.
186 Many of these residues are important for the helix-turn structure, while I434, L437, and
187 L438 are reported to regulate CD4-Lck interactions and endocytosis, suggesting that
188 these residues are under purifying selection due to their role in a multi-functional hub
189 (Kim et al., 2003; Sleckman et al., 1992). Intriguingly, the HIV-1 Nef protein interacts
190 with the IKRLL motif to promote CD4 downregulation (Kwon et al., 2020). Such host-
191 pathogen interactions often leave evidence of diversifying selection in host genes yet, in
192 this case, it is possible that HIV-1 Nef targets the IKRLL motif due to its functional
193 necessity, limiting the ability of CD4 to 'evolve around' Nef binding (Daugherty and
194 Malik, 2012; Elde et al., 2009).

195 Finally, C-terminal to the CQC clasp, mammalian CD4 contains a consensus
196 HRΦQK motif (mouse 448-452 in which Φ represents a large hydrophobic residue;
197 **Figure 1A**). This putative motif is not present in extant fish, reptiles, birds, or even the
198 marsupial CD4 orthologues sequenced to date. Yet, within the mammalian dataset, the
199 codons for H448 and R449 were found to be under purifying selection (**Figure 1C**). The
200 NMR solution structure of the CD4 ICD indicates that this region is unstructured within
201 human CD4, thus these residues are likely to be of functional importance (Kim et al.,
202 2003).

203 To summarize, our analyses highlight motifs in the ECD, TMD, and ICD of extant
204 eutherian CD4 that contain residues under purifying selection, indicating that they are
205 important for CD4 structure and/or function. We therefore proceeded to investigate how
206 residues within these motifs impact CD4 membrane domain localization, CD4-Lck

207 association, and pMHCII-responses.

208

209 **The GGXXG and CV+C motifs together impact membrane domain localization and**

210 **function**

211 We first tested the hypothesis that together the GGXXG and CV+C influence membrane

212 domain localization and function using previously established approaches (Fragoso et

213 al., 2003; Parrish et al., 2016). We included the CQC clasp because Lck has

214 myristylation and palmitoylation sites that could influence membrane domain localization

215 of CD4 when associated via the clasp (Ladygina et al., 2011). 58$\alpha^-\beta^-$ cells, which lack

216 expression of endogenous TCR and CD4, were generated expressing the 5c.c7 TCR,

217 which recognizes the 88-103 moth cytochrome c peptide (MCC) presented by mouse

218 MHCII I-E^k, along with CD4 (WT), a GGXXG to GVXXL mutant (TMD), a CV+C to SV+S

219 mutant (Palm), a CQC to SQS clasp mutant (Clasp), a double TMD+Palm (TP) mutant,

220 or a triple TMD+Palm+Clasp (TPC) mutant for functional analysis (**Figure**

221 **S2A**)(Letourneur and Malissen, 1989).

222 To measure membrane domain localization and Lck association we performed

223 sucrose gradient fractionation of detergent lysates. Proteins that float on the gradient

224 localize to detergent resistant membrane domains (DRMs) rich in membrane raft

225 components such as cholesterol and sphingolipids (Pike, 2006). The remaining proteins

226 localize to detergent soluble membrane domains (DSMs). CD4 molecules in each

227 fraction were immunoprecipitated (IP'd) onto polystyrene beads and stained with anti-

228 CD4, the cholera toxin B-subunit (CTxB) that binds GM1 gangliosides associated with

229 particular membrane rafts, and anti-Lck. Quantification of signal intensity was performed
230 by flow cytometry (**Figure S2B-S2D**).

231 We focused on the percent of CD4 signal in each fraction, relative to the total, to
232 account for any differences in the amount of CD4 between samples and independent
233 experiments (**Figure 2A**). Area under the curve (AUC) analysis of the DRMs and DSMs
234 revealed that Palm and Clasp mutant localization to the DRMs was lower than WT in
235 our sample size, while DRM localization for the TP and TPC mutants was further
236 reduced. These data indicate that the CQC and CV+C motifs, and particularly the
237 GGXXG plus CV+C together, mediate CD4 localization to the DRMs. The TP and TPC
238 showed slight increases in localization to the DSMs.

239 We next normalized the CTxB signal in each fraction to the CD4 signal in that
240 fraction to assess the amount of GM1 that co-IP'd with CD4 per fraction (**Figure 2B**).
241 AUC analysis revealed that the Clasp, TP, and TPC mutants within the DRMs had
242 reduced staining with CTxB, and the TP mutant had lower CTxB staining than the Palm
243 mutant. These data indicate that, for CD4 molecules within the DRM, the clasp
244 influences CD4 association with GM1-containing rafts and the GGXXG and CV+C
245 residues together do the same.

246 We also normalized the Lck signal in each fraction to the CD4 signal detected in
247 that fraction to analyze the amount of Lck that co-IP'd with CD4 per fraction (**Figure**
248 **2C**). The Palm, Clasp, TP and TPC signals were all greatly reduced for AUC analysis of
249 the DRM, indicating that both the CV+C and clasp motifs influence association with Lck
250 in the DRM. AUC analysis also showed that only the Clasp and TPC mutants had
251 reduced Lck association in the DSM, whereas the Palm and particularly the TP mutant

252 had increased association with Lck relative to the WT or T mutant alone. These data
253 indicate that the clasp controls if CD4 interacts with Lck, while the GGXXG and CV+C
254 motifs together influence the type of membrane domain in which Lck-associated CD4
255 molecules localize.

256 IL-2 production was measured in response to a titration of MCC peptide to
257 determine how these motifs influence pMHCII responses. If the frequency of CD4-Lck
258 interactions is the chief determinant for pMHCII responses then only the Clasp and TPC
259 mutants should reduce IL-2 production as the Palm and TP mutants interacted with Lck
260 in the DSM (Glaichenhaus et al., 1991; Stepanek et al., 2014). But, if CD4 association
261 with Lck in the DRMs is important, then the Palm and TP mutants would be expected to
262 have reduced IL-2 production. We observed a hierarchy of IL-2 production of WT > TMD
263 > Palm > Clasp ≥ TP ≥ TPC in response to a titration of MCC (**Figure 2D**), with AUC
264 analysis showing that the magnitude of the response across the peptide titration was
265 lower for each mutant cell line compared with the WT. The TP mutant also produced
266 less IL-2 than the Palm mutant. Finally, we observed the same hierarchy of IL-2
267 production in response to the lowest dose of MCC tested (41nM). These data suggest
268 the CV+C and GGXXG motifs that co-arose in evolution work together to enhance
269 pMHCII responses by localizing Lck-associated molecules in the DRM.

270

271 **The HXXR and HRΦQK motifs influence membrane domain localization and**
272 **association with Lck**

273 We next analyzed the HXXR and HRΦQK motifs which both contain core basic
274 residues under purifying selection. 58$\alpha^-\beta^-$ cells expressing a R426A mutant of the HXXR

275 motif, a R449A or a K452A mutant of the HRΦQK motif, or a R449A.K452A double
276 mutant (Dbl) were generated to explore how these residues work individually or in
277 combination. Each mutation reduced CD4 surface expression and CD4 signal in the
278 sucrose gradients relative to WT, suggesting that these residues influence CD4
279 expression or turnover (**Figure S3A**). While the CTxB staining was unremarkable, the
280 raw Lck signal was interesting when considered with CD4 levels because, while R426A
281 and R449A had lower Lck signal in the DRMs, the signal was elevated in the DSMs
282 compared with WT (**Figure S3B and S3C**). For these mutants we therefore calculated
283 the total AUC for the entire sucrose fractionation curve for raw Lck signal (**Figure S3D**),
284 which yielded a measure of the total Lck signal co-IP'd with the mutant CD4s that was
285 not different from the WT (Mean ± SEM: WT = 2628 ± 208; R429A = 2921 ± 377;
286 R449A = 2645 ± 430; K452A = 2385 ± 267; Dbl = 2155 ± 198; ns by One-way ANOVA
287 with Dunnett's posttest for mutants compared with WT), suggesting a higher overall
288 frequency of CD4-Lck association in the mutant cells.
289 Analysis of normalized sucrose gradient data revealed that R426A and R449A
290 reduced CD4 localization to the DRM, while K452A and the Dbl mutant had less of an
291 impact. The R426A mutant also increased localization to the DSM (**Figure 3A**). The
292 CTxB signal associated with CD4 in the DRMs or DSMs for these mutants was variable
293 (**Figure 3B**). Similarly, the Lck signal normalized to the CD4 signal for the DRMs was
294 unremarkable, but the R426A and R449A mutants were associated with more Lck
295 signal in the DSMs than the WT (**Figure 3C**). The K452 mutant had lower levels of
296 associated Lck than the R449A mutant in the DSM and the double mutant mirrored the
297 phenotype of the K452A mutant, indicating that the R449 and K452 residues are not

298 equivalent in the HRΦQK motifs. Together, the data indicate that the HxxR and HRΦQK
299 motifs influence CD4 membrane domain localization and association with Lck therein.
300 Here again, if the frequency of CD4-Lck association alone determines pMHCII-
301 responses then the mutants should have equivalent pMHCII-responses to the WT. But
302 AUC analysis showed that each mutant cell line produced less IL-2 than the WT (**Figure**
303 **3D**). These data are consistent with the idea that the membrane domain in which CD4
304 interacts with Lck is important for pMHCII responses.
305
306 **The intracellular helix interacts with Lck and attenuates pMHCII responses**
307 Focusing on the intracellular helix-turn structure of the ICD, we replaced residues 430-
308 442 with NGPGGNPGGNAGG to disrupt the chemical and structural nature of the helix-
309 turn region but maintain its length. We also combined this helix (H) mutant with the
310 clasp mutant (HC) to explore how they work together. The H and HC mutants had
311 elevated expression relative to the control, consistent with the helix regulating
312 endocytosis, and led to changes in CTxB and Lck staining (**Figure S4A-S4D**)(Sleckman
313 et al., 1992).
314 Normalized sucrose data showed that the percent of CD4 localized to the DRMs
315 and DSMs was not different for the H and HC mutants relative to the WT (**Fig 4A)**. The
316 HC mutant greatly reduced the normalized CTxB signal in the DRMs, while both H and
317 HC reduced the signal in the DSMs (**Figure 4B**). They also reduced the normalized Lck
318 signal associated with CD4 in the DRM and DSM relative to the wild-type (**Figure 4C**).
319 These data indicate that mutating the helix alone relieves association with Lck.

320 Given that the H and HC mutants failed to associate with Lck we expected both
321 mutants to significantly impair pMHCII-responses; yet, we found that both the H and HC
322 mutants increased the magnitude of the response across the peptide titration as well as
323 the sensitivity to the lowest peptide dose (**Figure 4D**). Also, consistent with prior reports
324 that the helix plays a role in endocytosis, the H and HC mutants did not show reduced
325 surface expression after 16hrs in these functional assays (**Figure S4E**). These data
326 show that without the helix, CD4-Lck association is not necessary for CD4 to enhance
327 pMHCII responses.
328
329 **The IKRLL interacts with Lck and attenuates pMHCII responses**
330 We next asked if specific residues within the helix were responsible for the H mutant
331 phenotype. Accordingly, we mutated the dileucine repeat (L437A+L438A = LL) within
332 the IKRLL motif because I/LXXLL motifs and dileucine repeats are known protein
333 interaction sites that could mediate interactions with a variety of partners; likewise, we
334 mutated the serine residues (S432A.S439A = SS) which could play a similar role upon
335 (de)phosphorylation. The SS mutant had slightly lower expression than the WT, and
336 both mutants showed variations in raw CTxB and Lck signal (**Figure S5A-S5D**). The
337 normalized sucrose gradient data revealed minor variations in the percent CD4 signal in
338 the DRMs and DSMs for the LL mutant and SS mutants relative to the WT (**Figure 5A**).
339 Likewise, within the DRM and DSMs there were no remarkable difference in CTxB
340 signal between the WT and mutants (**Figure 5B**). Finally, the LL mutant showed
341 reduced association with Lck in the DRMs and DSMs relative to the WT, which is
342 consistent with the IKRLL motif interacting with Lck (**Figure 5C**) (Kim et al., 2003;

343 Sleckman et al., 1992). In contrast, the SS mutant slightly increased interactions with
344 Lck in the DRM, compared to the WT, and greatly increased this interaction in DSM
345 which is consistent with regulating CD4-Lck interactions (Sleckman et al., 1992). These
346 data show that residues within the helix differentially regulate CD4 association with Lck.
347 If CD4 association with Lck is the chief determinant for pMHCII-responses then
348 the LL mutant should reduce IL-2 production by our 58$\alpha^-\beta^-$ cells in response to cognate
349 peptide while the SS mutant should enhance responses. However, we observed that
350 the LL mutation greatly increased both the magnitude of the response as well as the
351 sensitivity to a low peptide dose, while the SS did not differ from the WT (**Figure 5D**).
352 Both mutations slightly impaired pMHCII-induced endocytosis of CD4, consistent with
353 their previously repeated role in regulating this function (**Figure S5E**). Taken together
354 with the H and HC mutant phenotypes, we interpret the sum of the data to indicate that
355 the dominant function of the helix is to inhibit CD4's contribution to pMHCII responses
356 via the IKRLL motif.
357

358 **Coevolution of motifs in the ECD and ICD**

359 Because a number of the motifs considered above co-arose in mammals or eutherians
360 and, for the GGXXG and CV+C motifs, showed evidence of coordinated function, we
361 were motivated to explore if residues in our regions of interest showed evidence of
362 covariation. Constraints on protein function can lead to correlated mutations between
363 residues in a protein that provide further evidence of their functional importance and can
364 highlight networks of functional residues within a protein (Lockless and Ranganathan,
365 1999). We therefore used MISTIC2 to calculate the covariation between residues of

366 CD4 (Colell et al., 2018; Kowarsch et al., 2010). MISTIC2 quantifies correlations using
367 mutual information as a measure for how much information one random variable
368 provides about another, allowing for detection of coevolutionary relationships between
369 residues that are spatially distant and not just those that are in close proximity.

370 By analyzing the full data set we identified a network of interactions involving
371 residues within the helix-turn. Specifically, five pairs of coevolving residues were found
372 within the ICD helix-turn region (S432 – I434; S432 – S439; I434 – K441; L437 – S439;
373 L437 – K442), which may be relevant to the structure of this region, its function, or both
374 (**Figure 6**). Interestingly, G402 covaries with L438, suggesting a functional interaction
375 between the GGXXG and IKRLL motifs as well as a covariation network between the
376 TMD and ICD. Furthermore, H423 coevolved with I434 and S439, and R426 coevolved
377 with S439. Prior structural data suggests that the HXXR is unstructured and not
378 physically contacting the helix, suggesting instead a functional interplay between the
379 HxxR motif and the helix (Kim et al., 2003). Remarkably, we also found that P228 and
380 P281 of the nonpolar patch in the D3 domain of the ECD shows strong covariation with
381 residues in the ICD helix. Specifically, P228 coevolved with S432, I434, S439, and
382 K441, while P281 coevolved with S432, L438, and K442, providing evidence for the
383 evolution of intracellular regulation of an extracellular event. These data led us to
384 postulate that when the nonpolar patch arose, providing a mammalian CD4 ancestor
385 with extracellular interactions that enhance pMHCII responses, the increased signaling
386 was advantageous under some circumstances but detrimental in others; the conditions
387 were thus set for the evolution of a network of regulatory motifs that allowed for
388 finetuning of CD4's role in the relay of pMHCII-specific information from the outside to

389 the inside of a $CD4^+$ T cell.

390

391 **Evidence for a functional interplay between coevolving motifs**

392 To evaluate how the coevolution of these motifs impact pMHCII responses we assessed
393 IL-2 responses by $58\alpha^-\beta^-$ cells expressing CD4 molecules that have both an enhancing
394 motif and the inhibitor IKRLL motif mutated. We wanted to know if one motif would
395 dominate over the other, or if we would observe an interplay between the motifs.

396 We combined an ELXE mutant (P228E+F231E) of the PLXF motif in the D3
397 domain, which we reported to reduce $58\alpha^-\beta^-$ IL-2 responses, with the LL mutant to
398 assess if the helix inhibits contributions from the nonpolar patch (**Figure S6A**)
399 (Glassman et al., 2018). We also combined a previously described GKGVLIR to
400 GDGDSDS mutant in the D1 domain ($CD4^{\Delta bind}$), which kills CD4 binding to pMHCII and
401 pMHCII responses, to confirm that the LL mutant phenotype is dependent CD4-pMHCII
402 interactions (Glassman et al., 2016; Parrish et al., 2015). We found that the LL + ELXE
403 double mutant yielded similar results to the WT with regards to the magnitude and
404 sensitivity of the response to MCC, while the LL + Δbind double mutant completely
405 impaired responses (**Figure 7A and 7B**). Likewise, combining the TMD and Palm
406 mutants with the dileucine mutant (TP+LL) reduced the magnitude of IL-2 production
407 across a MCC titration, and at the lowest peptide dose, relative to the LL mutant such
408 that the response was indistinguishable from the WT (**Figure 7A, 7B, and S6C**).
409 Similarly, combining the R426A or Clasp mutants with the LL mutation generated similar
410 responses to the WT.

411 These data indicate that the LL mutation does not enhance pMHCII responses in

412 the absence of CD4-pMHCII interactions. Rather, the ICD helix can counter-balance the
413 impact of ECD, TMD, and ICD enhancing motifs. Our results with the LL + Clasp double
414 mutant are also consistent with our HC mutant phenotype (**Figure 4D**), supporting the
415 idea that CD4-Lck interactions are not necessary for pMHCII responses in the absence
416 of the inhibitory function of the IKRLL motif. Together, the data provide evidence for a
417 functional interplay between motifs that are coevolving with the inhibitory helix in the
418 ICD to regulate events occurring outside and inside the cell.
419

420 **DISCUSSION**

421 Biology has been experimenting with the form and function of CD4 *in vivo* for ~435
422 million years in a variety of jawed vertebrates. The results of these experiments, as
423 deconvolved by our computational reconstruction of CD4 evolution, thus provide
424 insights into the residues or motifs that are essential for CD4 to confer fitness to jawed
425 vertebrates within a particular clade or species in a way that cannot be readily replicated
426 through reductionist studies of a single species or genetically engineered model thereof.
427 Coupling our computational results with biochemical and functional analyses then
428 provides us with mechanistic insights into why these residues or motifs are important for
429 fitness. Together, our multidisciplinary approach provides evidence that eutherian CD4
430 evolved an epistatic network of function-regulating motifs wherein several of the
431 enhancing motifs, including the nonpolar patch in the ECD, coevolved with an inhibitory
432 helix in the ICD. The results provide insights into the evolution and function of the multi-
433 subunit receptor system that relays pMHCII-specific information from the outside to the
434 inside of $CD4^+$ T cells.

435 The data presented here advance our understanding of eutherian CD4 by
436 showing that the TMD and ICD evolved motifs that influence where Lck-associated CD4
437 molecules localize, and how well they drive pMHCII responses. As in prior work, we
438 found that the CQC and CV+C motifs individually influence CD4 localization to DRMs
439 (Fragoso et al., 2003). We extend this knowledge by showing that the CQC clasp, but
440 not the CV+C motif alone, influences if CD4 molecules within DRMs inhabit GM1-
441 containing rafts. Furthermore, we show that together the GGXXG and CV+C motifs
442 localize Lck-associated CD4 molecules to DRMs and help localize CD4 molecules
443 within DRMs to GM1-containing raft. We also extend prior insights about the HXXR
444 motif by showing that it minimizes accumulation of Lck-associated CD4 molecules to
445 DSMs, enhances CD4 expression, and enhances pMHCII responses (Popik and Alce,
446 2004). Similar results are presented for the HRΦQK motif. One interpretation of these
447 data, which fits within the conceptual framework for the role of membrane rafts in cell
448 signaling, is that eutherian CD4 is under selective pressure to regulate the membrane
449 domain in which CD4 interacts with Lck because localization to membrane rafts, not
450 simply interacting with Lck, is a key determinant for pMHCII responses. But this
451 interpretation is only accurate when the IKRLL motif of the intracellular helix is intact, as
452 without this motif CD4-Lck interactions are not necessary. We therefore conclude that
453 the inhibitory function of the helix counterbalances the activity of the enhancing motifs to
454 both prevent those CD4 molecules that are not associated with Lck from contributing to
455 signaling and enforce a requirement that CD4-Lck interactions occur in membrane rafts
456 for optimal pMHCII responses.
457 Our results also offer clues as to why and how eutherian CD4 evolved such

458 elaborate control over CD4 activity. We find that the residue equivalent to P281 is
459 present in all eutherian CD4 molecules but absent in the other amniotes we analyzed.
460 This residue is a constituent of the nonpolar patch in the D3 domain of CD4 that
461 contributes to the formation and stability of TCR-CD3-pMHCII-CD4 assemblies. We
462 posit that the acquisition of P281 may have been a large-effect mutation, as described
463 for other proteins, that conferred upon CD4 enhanced function and set the conditions for
464 the rapid acquisition of additional mutations to balance the impact of the nonpolar patch
465 (Whitney et al., 2016). A *de novo* insertion(s) between the TMD and the clasp of the
466 amniote MRCA ICD (node 1), which likely explains the size difference between the fish
467 and amniote CD4 ICD, would provide evolution with a substrate to explore novel
468 sequences and add function to a previously non-functional region through the selection
469 of beneficial mutations (Davey et al., 2015; Van Roey et al., 2012). In this scenario,
470 individual motifs that impact the functional success of one another become dependent
471 upon each other, as the fitness cost to change any single motif would be too large. This
472 scenario is supported by the conservation of motifs we observe with residues under
473 purifying selection, the covariation network we identified between the nonpolar patch in
474 the ECD with the ICD helix, and our experimental analysis. It is also in line with the
475 model of residue entrenchment described elsewhere (Shah et al., 2015).
476 The outcome for eutherian CD4 is the accumulation of coevolving motifs with
477 switch-like capabilities to finetune the relay of pMHCII-specific information across the
478 cell membrane. For example, the function of the CV+C motif is likely to be dynamic as
479 palmitoylation is reversible (Ladygina et al., 2011). Furthermore, GXXXG motifs mediate
480 interactions with other proteins, or with cholesterol which can modulate protein-protein

481 interactions or induce allosteric changes (Song et al., 2014; Teese and Langosch,
482 2015). As cholesterol concentrations differ in distinct membrane domains, and can
483 change upon TCR signaling, the GGXXG motif in eutherian CD4 may assume different
484 functions depending on the membrane domain in which it is located or the activation or
485 differentiation state of the T cell (Fessler, 2016). Of note, palmitate can interact with
486 protein-bound cholesterol (PDB 4IB4), making it plausible that coupling of the
487 palmitoylation state of the CV+C motif with a cholesterol-bound state of the GGXXG
488 motif could further influence CD4 membrane domain localization and function. The
489 HXXR and HRΦQK motifs are also likely to have dynamic functions as both are rich in
490 basic residues that could form reversible electrostatic interactions with acidic
491 headgroups of particular lipid species, serve as docking sites for low-affinity interactions
492 with proteins or second messenger ligands, or both (Aivazian and Stern, 2000; Davey et
493 al., 2015). For example, HXXR motifs are common in the P2Y family of GPCRs and
494 related cysteinyl-leukotriene receptors, where they mediate ligand binding (Parravicini et
495 al., 2010). Moreover, a RXXK motif in Slp-76 has been described as a nonproline-based
496 binding site for the SH3 domains of Gads and Grb2; the histidine residue of the HRΦQK
497 motif, which is under purifying selection, would not favor interactions with Gads and
498 Grb2 according to that study, but it may define binding specificity for other SH3-domain
499 bearing interaction partners (Berry et al., 2002). Of note, changes in membrane
500 proximal cytoplasmic pH upon T cell activation could alter the charge of the histidine
501 imidazole side-chain and favor particular interactions, perhaps to Gads or Grb2, under
502 very specific conditions giving this motif switch like function. Finally, the helix appears to
503 be a multifunctional hub, with basic residues that could mediate interactions with the

504 membrane; serines that could be phosphorylated to repel the ICD from the membrane,
505 stabilize interactions with other binding partners, or disrupt binding to Lck; and an IKRLL
506 protein interaction motif that interacts with Lck but is likely to have other partners
507 involved in its inhibitory function (Aivazian and Stern, 2000; Garron et al., 2008; Kelly et
508 al., 2008; Sleckman et al., 1992; Yeung and Grinstein, 2007). Such an assembly of
509 switch functions could thus provide tunability to CD4 function. Identifying interaction
510 partners for these motifs, and deconvolving how they work in concert, will be the subject
511 of future work.

512 In closing, the eutherian immune system achieves balance in its response to
513 pMHCII not only through the interplay of helper and regulatory $CD4^+$ T cells, or
514 costimulatory and coinhibitory molecules, but also through finetuning CD4 function with
515 enhancing and inhibitory motifs. Our findings are consistent with the idea that CD4
516 arose to enable more elaborate pMHCII responses than could be achieved otherwise
517 (Venkatesh et al., 2014); indeed, it is interesting that the timing of key stepwise events
518 in the evolution of *FoxP3*, or even helper and memory T cell functions, exhibit parallels
519 with CD4 (Andersen et al., 2012; Lane et al., 2010). We expect that the coevolution of
520 regulatory elements outside and inside of a cell, which to our knowledge is first
521 evidenced here, will emerge as a theme for other transmembrane receptors tasked with
522 relaying information to its intracellular signaling machinery. Finally, given that efforts to
523 engineer synthetic receptors for therapeutic purposes are showing promise, most
524 notably for chimeric antigen receptor (CAR)-T cell therapy, these results have
525 translational value. The archetypal CAR design was conceived when it was understood
526 that receptor bind ligands outside of a cell and generate signals inside the cell, but

527 knowledge of the intermediate steps was lacking (Eshhar et al., 1993). Problems with
528 sensitivity and side effects now suggest that the absence of mechanisms to mediate or
529 regulate the relay of information across the membrane have a fitness cost for this form
530 of CAR (Labanieh et al., 2018). We therefore think that biomimetic designs,
531 incorporating strategies refined over ~435 million years of iterative testing in a variety of
532 vertebrates, will ultimately lead to more sensitive and reliable synthetic receptors
533 (Kobayashi et al., 2020). Such biomimetic engineering will require a doubling down on
534 basic research efforts to elucidate the evolutionary blueprint for key immune receptors.
535

536 **MATERIALS AND METHODS**

537

538 **Evolutionary analyses**

539 Available CD4 orthologs were downloaded from GenBank. The criteria used for
540 including orthologous CD4 sequences in our analyses were that they have a domain
541 structure consisting of four extracellular Ig domains followed by a TMD and a C-terminal
542 ICD. Sequences that were shorter, contained frameshift mutations, etc…, were
543 excluded from the analysis. Teleost fish were considered to be the oldest living species
544 that contain a CD4 molecule in this study as a CD4 ortholog was not identified in
545 cartilaginous fish (Venkatesh et al., 2014). The final dataset contained 99 unique CD4
546 orthologs, ranging from teleost fish to human. These (putative) coding sequences were
547 translated to amino-acids and aligned using MAFFT (Katoh et al., 2002). For codon-
548 based analyses, the aligned amino acid sequences were back translated to nucleotides
549 to maintain codons. The multiple sequence alignments were further processed to
550 remove all insertions (indels) relative to the mouse CD4 sequence (NM_013488.3) to
551 maintain consistent numbering of sites. The 5' and 3' regions of the CD4 molecules
552 were not consistently aligned due to different start codon usage or extensions of the
553 ICD, respectively. The alignment was edited to start at the codon (AAG) coding for K48
554 within mouse CD4. The alignment that includes all 99 CD4 sequences ends at the last
555 cysteine residue that makes up the CQC clasp. For the mammalian only dataset, the
556 alignment terminates at the mouse CD4 stop codon.

557 FastTree was used to estimate maximum likelihood trees (Price et al., 2010). For
558 amino acid-based trees, the Jones-Taylor-Thornton (JTT) model of evolution was

559 selected, while the General Time-Reversible (GTR) model was used for nucleotide trees
560 (Jones et al., 1992; Waddell and Steel, 1997). In all conditions, a discrete gamma model
561 with 20 rate categories was used. A reduced representation tree (based on the amino
562 acid alignment) is shown in Figure 1A. We used the same JTT model to estimate the
563 marginal reconstructions of nodes indicated in Figure 1A. Phylogenetic trees and logo
564 plots were visualized in Geneious and further edited using Adobe Illustrator.
565 Ancestral sequences were estimated using GRASP (Foley et al., 2020).
566 Codon-based analysis of selection was performed using the hypothesis testing
567 fixed effects likelihood (FEL) model as implemented within the phylogenies (HyPhy)
568 package (version 2.5.14) (MP) (Kosakovsky Pond et al., 2020; Pond et al., 2005; Pond
569 and Frost, 2005; Weaver et al., 2018). The back-translated codon-based alignments
570 described above were used for these analyses. FEL uses likelihood ratio tests (LRTs) to
571 assess a better fit of codons that allowed selection ($P < 0.1$). When calculating values
572 for all CD4 orthologs included in the initial phylogenetic analysis we analyzed and
573 identified the sequences within the mammalian clade as the foreground branches on
574 which to test for evolutionary selection in order to maximize statistical power.
575 Covariation between protein residues was calculated using the MISTIC2 server.
576 We calculated 4 different covariation methods (MIp, mFDCA, plmDCA, and
577 gaussianDCA)(Colell et al., 2018). Protein conservation scores were calculated based
578 on the protein alignment using the ConSurf Server (Ashkenazy et al., 2016; Ashkenazy
579 et al., 2010). ConSurf conservation scores are normalized, so that the average score for
580 all residues is zero, with a standard deviation of one. The lower the score, the more
581 conserved the protein position. For the purpose of this study, residues were considered

582 to covary if the MI was larger than 4 and both residues had a ConSurf conservation
583 score lower than -0.5. Also, pairs with an MI larger than 8 were considered to covary if
584 the conservation score was below -0.3. Using these criteria, we selected 0.5% of all
585 possible pairs as recommended (Buslje et al., 2009; Colell et al., 2018).
586 Raw data, including alignments and phylogenetic trees, associated with figures 1,
587 S1, and 6 are available on Dryad (https://doi.org/10.5061/dryad.59zw3r26z)..
588
589 **Cell lines**
590 58$\alpha^-\beta^-$ T cell hybridoma lines were generated by retroviral transduction and maintained
591 in culture by standard techniques as previously described (Glassman et al., 2018;
592 Letourneur and Malissen, 1989). In brief, one day after transduction the cells were
593 cultured in 5µg/ml puromycin(Invivogen) and 5µg/ml zeocin(ThermoFisher Scientific) in
594 RPMI 1640 (Gibco) supplemented with 5% FBS (Atlanta Biologicals or Omega
595 Scientific), Penicillin-Streptomycin-Glutamine (Cytiva), and 50µM βME (ThermoFisher
596 Scientific). The next day drug concentrations were increased to 10µg/ml puromycin
597 (Invivogen) and 100µg/ml zeocin (ThermoFisher Scientific) in 10ml in a T25 flask.
598 Aliquot of $1x10^7$ cells were frozen at days 5, 7, and 9. Cells were thawed from the day 5
599 freeze and cultured for 3 days in 10µg/ml puromycin and 100µg/ml zeocin, and
600 maintained below $1x10^6$ cells/ml to use in the functional assays. Cells used in the
601 functional assays were grown to $0.8x10^6$ cells/ml density and replicates of three
602 functional assays were performed every other day. If cells exceeded $1x10^6$ cells/ml at
603 any point in the process they were discarded as they lose reactivity at high cell densities
604 and a new set of vials was thawed. Typically, two independent WT and mutant pairs

605 were generated for any given mutant and tested for IL-2 to gain further confidence in a
606 response phenotype. When cells lines are presented together in a graph, that indicates
607 that the cell lines (WT and mutants sets) were generated at the same time from the
608 same parental cell stock.

609 Given the number of mutant CD4 cell lines generated and handled in this study,
610 the identity of the transduced CD4 gene was verified by PCR sequencing at the
611 conclusion of three independent functional assays. 58$\alpha^-\beta^-$ T cell hybridomas were lysed
612 using DirectPCR Tail Lysis Buffer (Viagen Biotech) with proteinase K (SIGMA) for 2
613 hours at 65˚C. Cells were then heated at 95˚C for ten minutes. Cell debris was pelleted,
614 and the supernatants were saved. CD4 was amplified by PCR using Q5 DNA Hot Start
615 Polymerase (New England BioLabs) in 0.2μM dNTP, 0.2μM primer concentration, Q5
616 Reaction buffer, and water. CD4 was amplified using the following primers:
617 5' primer: acggaattccgctcgagcgccaccatggtgcgagccatctctctcttagg
618 3' primer: ctagcaagcttgtcgactcaagatcttcattagatgagattatggctcttctgc
619 Product were purified using SpinSmart Nucleic Acid Purification Columns (Thomas
620 Scientific) and sent to Eton Bioscience for sequencing with the following 5' CD4 primer:
621 gtctctgaggagcagaag.

622 The I-E^{k+} M12 cells used as antigen presenting cells (APCs) were previously
623 reported (Glassman et al., 2018). Cells were cultured in RPMI 1640 (Gibco)
624 supplemented with 5% FBS (Atlanta Biologicals or Omega Scientific), Pencillin-
625 Streptomycin-Glutamine (Cytiva), 50μM βME (ThermoFisher Scientific), and 5μg/mL
626 Puromycin (Invivogen) and 50μg/ml Zeocin (ThermoFisher).
627

628 **Flow cytometry**

629 Cell surface expression of CD4 and TCR-CD3 complexes were measured by flow
630 cytometry. In brief, cells were stained for 30 minutes at 4ºC in FACS buffer (PBS, 2%
631 FBS, and 0.02% sodium azide) using anti-CD4 (clone GK1.5, eFluor 450 conjugate,
632 ThermoFisher Scientific), anti-TCRα (anti-Vα11, clone RR8-1, APC conjugate,
633 ThermoFisher Scientific), anti-CD3ε (145-2C11, ThermoFisher Scientific) and GFP was
634 detected as a measure of the TCRβ-GFP subunit. Analysis was performed on a Canto II
635 or LSRII (BD Biosciences) at the Flow Cytometry Shared Resource at the University of
636 Arizona. Flow cytometry data were analyzed with FlowJo Version 9 software (Becton,
637 Dickinson & Company).

638

639 **Functional assays**

640 IL-2 production was measured to quantify pMHCII responses. $5x10^4$ transduced $58\alpha^-\beta^-$
641 T cell hybridomas were cocultured with $1X10^5$ transduced I-E^{k+} M12 cells in a 96 well
642 round bottom plate in the presence of titrating amounts of MCC 88-103 peptide
643 (purchased from 21st Century Biochemicals at >95% purity) starting at 30μM MCC and a
644 1:3 titration (Glassman et al., 2018). The supernatants were collected and assayed for
645 IL-2 concentration by ELISA after 16 hours of co-culture at 37ºC. Anti-mouse IL-2 (clone
646 JES6-1A12, BioLegend) antibody was used to capture IL-2 from the supernatants, and
647 biotin anti-mouse IL-2 (clone JES6-5H4, BioLegend) antibody was used as the
648 secondary antibody. Streptavidin-HRP (BioLegend) and TMB substrate (BioLegend)
649 were also used.

650 To assess engagement-induced endocytosis, CD4 surface levels were measured

651 by flow cytometry 16hrs after coculture with APCs and peptide as described above for
652 IL-2 quantification. 96 well plates containing cells were washed with ice cold FACs
653 buffer (PBS, 2%FBS, 0.02% sodium azide), transferred to ice, and Fc receptors were
654 blocked with Fc block mAb clone 2.4G2 for 15 minutes at 4˚C prior to surface staining
655 for 30 minutes at 4˚C with anti-CD4 (clone GK1.5 EF450, Invitrogen) and anti-Vβ3 TCR
656 clone (clone KJ25, BD Pharmingen) antibodies. Cells were washed with FACS buffer
657 prior to analysis on a LSRII (BD Biosciences) at the Flow Cytometry Shared Resource
658 at the University of Arizona. Flow cytometry data were analyzed with FlowJo Version 10
659 software (Becton, Dickinson & Company). The average of the geometric mean of the
660 TCR or CD4 signal was taken for the triplicate of the post 58$\alpha^-\beta^-$ cells cocultured with
661 M12 I-E^{k+} cells at 0μM MCC concentration. Each value of the raw gMFI of TCR or CD4
662 for cells cultured at 10μM MCC was subtracted from the average gMFI at 0μM. The
663 values show the change of gMFI from 0μM to 10μM.
664
665 **Sucrose gradient analysis**
666 Membrane fractionation by sucrose gradient was performed similarly to previously
667 described methods (Hur et al., 2003; Parrish et al., 2016). For cell lysis, $6x10^7$ 58$\alpha^-\beta^-$ T
668 cell hybridomas were harvested and washed 2x using TNE buffer (Tris, NaCl, EDTA).
669 Cells were lysed on ice in 1% Triton-X detergent in TNE in a total volume of 1mL for 10
670 minutes and then dounce homogenized 10x. The homogenized lysate was transferred
671 to 14 x 95mm Ultraclear Ultra Centrifuge tubes (Beckman). The dounce homogenizer
672 was rinsed with 1.6mL of the 1% lysis buffer, which was then was added to the
673 Ultracentrifuge tube. 2.5mL of 80% sucrose was added to the centrifuge tube with

674 lysate, and mixed well. Gently, 5mL of 30% sucrose was added to the centrifuge tubes,
675 creating a 30% sucrose layer above the ~40% sucrose/lysate mixture. Then, 3mL of 5%
676 sucrose was added gently to the centrifuge tube, creating another layer. The centrifuge
677 tubes were spun 18 hours at 4˚C in a SW40Ti rotor at 36,000rpm.

678 Analysis of membrane fractions was performed via flow-based fluorophore-linked
679 immunosorbent assay (FFLISA) as previously described (Parrish et al., 2016). In brief,
680 88µL of Streptavidin Microspheres 6.0µm (Polysciences) were coated overnight at 4ºC
681 with 8µg of biotinylated anti-CD4 antibody (clone RM4-4, BioLegend). Prior to
682 immunoprecipitation (IP), beads were washed with 10mL of FACS buffer (1x PBS,
683 2%FBS, 0.02% sodium azide) and resuspended in 3.5mL of 0.1% Triton-X 100 lysis
684 wash buffer in TNE. For each cell line lysed, 10 FACS tubes were prepared with 50uL of
685 the washed RM4-4 coated beads. Upon completion of the spin, 500µL was carefully
686 taken off the top of the centrifuge tubes and discarded. Following this, 1mL was
687 extracted from the top of the tube, carefully as to not disrupt the gradient, and added to
688 a FACS tube with coated beads and capped. This was repeated for 10 individual
689 fractions in separate FACS tubes. Following the extraction, lysates were incubated with
690 the beads for 90 minutes, inverting the tubes to mix every 15 minutes.

691 Following the immunoprecipitation, FACS tubes were washed 3x using 0.1%
692 Triton-X lysis wash buffer in TNE. Tubes were then stained using 1µL anti-CD4 (APC
693 conjugate; clone GK1.5, BioLegend), 1.5µL anti-Lck (PE conjugate, clone 3A5, Santa
694 Cruz Biotechnology), and 1µL CTxB (AF 488 conjugate, ThermoFisher Scientific,
695 resuspended as per manufacturers instructions) for 45 minutes at 4ºC. Following the
696 stain, tubes were washed using 0.1% Triton-X lysis wash buffer in TNE. Analysis of

697 beads were performed on a LSRII (BD Biosciences) at the Flow Cytometry Shared
698 Resource at the University of Arizona. 10^4 events were collected per sample. Flow
699 cytometry data were analyzed with FlowJo Version 10 software (Becton, Dickinson &
700 Company).

701 For FFLISA analysis, raw gMFI values for fraction 1 were subtracted from the
702 rest of the fractions to account for background, such that the gMFI of fraction 1 is 0. To
703 normalize the data, the percentage of CD4 within any given fraction (fx) relative to the
704 total CD4 gMFI (CD4 signal % of total) was calculated by dividing the gMFI signal in a
705 given fraction (fx) by the sum of the total CD4 gMFI signal [sum(f1:f10)CD4 gMFI] and
706 multiplying by 100 [e.g. fx % of total = fx CD4 gMFI / Sum(f1:f10)CD4 gMFI x 100]. To
707 normalize the CTxB and Lck signal in any given fraction relative to the CD4 signal in
708 that same fraction (CTxB or Lck normalized to CD4) the gMFI of CTxB or Lck in fx was
709 divided by the CD4 gMFI of that fx and then multiplied by the percentage of CD4 within
710 fx (e.g. Normalized fx Lck = fx Lck gMFI/fx CD4 gMFI x fx CD4% of total CD4 gMFI).
711 Area under the curve (AUC) analysis was performed with GraphPad Prism 9 for
712 fractions 1-6 to determine the AUC for the detergent resistant membrane (DRM)
713 domains due to their floating phenotypes, and for fractions 6-10 to determine the AUC
714 for the detergent soluble membrane (DSM) domains.

715

716 **Statistical Analysis**

717 Statistical analyses of sucrose gradient and functional assays were performed with
718 GraphPad Prism 9 software as indicated in the figure legends. For each functional
719 assay (IL-2 production and CD4 endocytosis), each individual experiment (biological

720 replicate) was performed with triplicate analysis (technical replicates) and each
721 experiment was repeated three times (3 biological replicates). For sucrose gradient
722 analysis, 10^4 beads were collected by flow cytometry in each experiment (technical
723 replicates) and each experiment was performed three times (biological replicates).
724 Three biological replicates were chosen for each analysis as per convention, and no
725 power calculations were determined. One-Way ANOVA were performed with a
726 Dunnett's posttest when all mutants tested in an experiment were compared to a control
727 sample (e.g. WT). Sidak's posttest were applied when comparing between two specific
728 samples. These posttests were chosen based on Prism recommendations.
729
730 **Constructs**
731 The following sequence for 5cc7α was subcloned into the pZ4 zeocin-resistance MSCV
732 vector (MCS-IRES-Zeo) resistance through 5'XhoI and 3'EcoRI:
733
734 cctcgagcgccaccatggagaggaacctgggagctgtgctggggattctgtgggtgcagatttgctgggtgagaggagat
735 caggtggagcagagtccttcagccctgagcctccacgagggaaccggttctgctctgagatgcaattttactaccaccatg
736 agggctgtgcagtggttccgaaagaattccaggggcagcctcatcaatctgttctacttggcttcaggaacaaaggagaat
737 gggaggctaaagtcagcatttgattctaaggagcgctacagcaccctgcacatcagggatgcccagctggaggactcag
738 gcacttacttctgtgctgctgaggcttccaataccaacaaagtcgtctttggaacagggaccagattacaagtattaccaaa
739 catccagaacccagaacctgctgtgtaccagttaaaagatcctcggtctcaggacagcaccctctgcctgttcaccgacttt
740 gactcccaaatcaatgtgccgaaaaccatggaatctggaacgttcatcactgacaaaactgtgctggacatgaaagctat
741 ggattccaagagcaatggggccattgcctggagcaaccagacaagcttcacctgccaagatatcttcaaagagaccaa
742 cgccacctaccccagttcagacgttccctgtgatgccacgttgaccgagaaaagctttgaaacagatatgaacctaaacttt
743 caaaacctgtcagttatgggactccgaatcctcctgctgaaagtagcgggatttaacctgctcatgacgctgaggctgtggt
744 ccagtgcggccgcatgataaagatctggatcctgactgagtgagaattc
745
746 The following sequence for 5cc7βG was subcloned into the pP2 puromycin-resistance
747 MSCV vector (MCS-IRES-puro) resistance through 5'XhoI and 3'BamHI:
748
749 actcgagcgccaccatggctacaaggctcctctgttacacagtactttgtctcctgggtgcaagaattttgaattcaaaagtc
750 attcagactccaagatatctggtgaaagggcaaggacaaaaagcaaagatgaggtgtatccctgaaaagggacatcca
751 gttgtattctggtatcaacaaaataagaacaatgagtttaaatttttgattaactttcagaatcaagaagttcttcagcaaatag
752 acatgactgaaaaacgattctctgctgagtgtccttcaaactcaccttgcagcctagaaattcagtcctctgaggcaggaga
753 ctcagcactgtacctctgtgccagcagtctgaacaatgcaaactccgactacaccttcggctcagggaccaggcttttggta
754 atagaggatctgagaaatgtgactccacccaaggtctccttgtttgagccatcaaaagcagagattgcaaacaaacaaa
755 aggctaccctcgtgtgcttggccaggggcttcttccctgaccacgtggagctgagctggtgggtgaatggcaaggaggtcc
756 acagtggggtcagcacggaccctcaggcctacaaggagagcaattatagccactgcctgagcagccgcctgagggtct

757 ctgctaccttctggcacaatcctcgcaaccacttccgctgccaagtgcagttccatgggctttcagaggaggacaagtggc
758 cagagggctcacccaaacctgtcacacagaacatcagtgcagaggcctggggccgagcagactgtgggattacctca
759 gcatcctatcaacaaggggtcttgtctgccaccatcctctatgagatcctgctagggaaagccaccctgtatgctgtgcttgtc
760 agtacactggtggtgatggctatggtcaaaagaaagaattccgcggccgcaggtggaggcggatcaggtggcggtgga
761 agtggaggtggtggatctatggtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacg
762 gcgacgtaaacggccacaagttcagcgtgtccggcgagggcgagggcgatgccacctacggcaagctgaccctgaa
763 gttcatctgcaccaccggcaagctgcccgtgccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgcttcag
764 ccgctaccccgaccacatgaagcagcacgacttcttcaagtccgccatgcccgaaggctacgtccaggagcgcaccat
765 cttcttcaaggacgacggcaactacaagacccgcgccgaggtgaagttcgagggcgacaccctggtgaaccgcatcga
766 gctgaagggcatcgacttcaaggaggacggcaacatcctggggcacaagctggagtacaactacaacagccacaac
767 gtctatatcatggccgacaagcagaagaacggcatcaaggtgaacttcaagatccgccacaacatcgaggacggcag
768 cgtgcagctcgccgaccactaccagcagaacacccccatcggcgacggccccgtgctgctgcccgacaaccactacct
769 gagcacccagtccaagctgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgccg
770 ccgggatcactctcggcatggacgagctgtacaagtaatgaggatcctga

771 Full-length CD3δ, ε, γ, and ζ were encoded on a poly-cistronic construct as previously
772 described (Holst et al., 2008; Kuhns and Davis, 2007)

773 CD4 WT, CD4 TMD, CD4 Palm, CD4 Clasp, CD4 TP, CD4 TPC were cloned by
774 conventional molecular biology techniques, including PCR-based mutagenesis where
775 needed, into pUC18 via 5'EcoRI and 3'HindIII. After sequence verification they were
776 subcloned into pP2 puromycin-resistance MSCV vector (MCS-IRES-puro) vectors via
777 5'XhoI and 3'BglII, midi-prepped (Qiagen), and sequence verified. The sequences for
778 these constructs are as follows. Mutated codons are shown in bold uppercase letters,
779 while the motif under interrogation is shown in bold and underlined (any non-mutated
780 codons within the motif are bold lowercase letters and underlined):

781 **CD4 WT:**
782 acggaattccgctcgagcgccaccatggtgcgagccatctctcttaggcgcttgctgctgctgctgcagctgtcacaac
783 tcctagctgtcactcaagggaagacgctggtgctggggaaggaaggggaatcagcagaactgccctgcgagagttccc
784 agaagaagatcacagtcttcacctggaagttctctgaccagaggaagattctggggcagcatggcaaaggtgtattaatta
785 gaggaggttcgccttcgcagtttgatcgttttgattccaaaaaaggggcatgggagaaaggatcgtttcctctcatcatcaat
786 aaacttaagatggaagactctcagacttatatctgtgagctggagaacaggaaagaggaggtggagttgtgggtgttcaa
787 agtgaccttcagtccgggtaccagcctgttgcaagggcagagcctgaccctgaccttggatagcaactctaaggtctctaa
788 ccccttgacagagtgcaaacacaaaaagggtaaagttgtcagtggttccaaagttctctccatgtccaacctaagggttca
789 ggacagcgacttctggaactgcaccgtgaccctggaccagaaaaagaactggttcggcatgacactctcagtgctgggtt
790 ttcagagcacagctatcacggcctataagagtgagggagagtcagcggagttctccttcccactcaactttgcagaggaa
791 aacgggtggggagagctgatgtggaaggcagagaaggattctttcttccagccctggatctccttctccataaagaacaa
792 agaggtgtccgtacaaaagtccaccaaagacctcaagctccagctgaaggaaacgctcccactcaccctcaagatacc
793 ccaggtctcgcttcagtttgctggttctggcaacctgactctgactctggacaaagggacactgcatcaggaagtgaacctg
794 gtggtgatgaaagtggctcagctcaacaatactttgacctgtgaggtgatgggacctacctctcccaagatgagactgacc
795 ctgaagcaggagaaccaggaggccagggtctctgaggagcagaaagtagttcaagtggtggcccctgagacagggct
796 gtggcagtgtctactgagtgaaggtgataaggtcaagatggactccaggatccaggttttatccagaggggtgaaccaga
797 cagtgttcctggcttgcgtgctgggtggctccttcggctttctgggtttccttgggctctgcatcctctgctgtgtcaggtgccggc
798 accaacagcgccaggcagcacgaatgtctcagatcaagaggctcctcagtgagaagaagacctgccagtgcccccac
799 cggatgcagaagagccataatctcatctaatgaagatcttgagtcgacaagcttgctag

800

801 **CD4 TMD:**

802 acggaattccgctcgagcgccaccatggtgcgagccatctctcttaggcgcttgctgctgctgctgcagctgtcacaac
803 tcctagctgtcactcaagggaagacgctggtgctggggaaggaaggggaatcagcagaactgccctgcgagagttccc
804 agaagaagatcacagtcttcacctggaagttctctgaccagaggaagattctggggcagcatggcaaaggtgtattaatta
805 gaggaggttcgccttcgcagtttgatcgttttgattccaaaaaaggggcatgggagaaaggatcgtttcctctcatcatcaat
806 aaacttaagatggaagactctcagacttatatctgtgagctggagaacaggaaagaggaggtggagttgtgggtgttcaa
807 agtgaccttcagtccgggtaccagcctgttgcaagggcagagcctgaccctgaccttggatagcaactctaaggtctctaa
808 ccccttgacagagtgcaaacacaaaaagggtaaagttgtcagtggttccaaagttctctccatgtccaacctaagggttca
809 ggacagcgacttctggaactgcaccgtgaccctggaccagaaaaagaactggttcggcatgacactctcagtgctgggtt
810 ttcagagcacagctatcacggcctataagagtgagggagagtcagcggagttctccttcccactcaactttgcagaggaa
811 aacgggtggggagagctgatgtggaaggcagagaaggattctttcttccagccctggatctccttctccataaagaacaa
812 agaggtgtccgtacaaaagtccaccaaagacctcaagctccagctgaaggaaacgctcccactcaccctcaagatacc
813 ccaggtctcgcttcagtttgctggttctggcaacctgactctgactctggacaaagggacactgcatcaggaagtgaacctg
814 gtggtgatgaaagtggctcagctcaacaatactttgacctgtgaggtgatgggacctacctctcccaagatgagactgacc
815 ctgaagcaggagaaccaggaggccagggtctctgaggagcagaaagtagttcaagtggtggcccctgagacagggct
816 gtggcagtgtctactgagtgaaggtgataaggtcaagatggactccaggatccaggttttatccagaggggtgaaccaga
817 cagtgttcctggcttgcgtgctgggt**GTGtccttcCTC**tttctgggtttccttgggctctgcatcctctgctgtgtcaggtgccg
818 gcaccaacagcgccaggcagcacgaatgtctcagatcaagaggctcctcagtgagaagaagacctgccagtgccccc
819 accggatgcagaagagccataatctcatctaatgaagatcttgagtcgacaagcttgctag

820

821 **CD4 Palm:**

822 acggaattccgctcgagcgccaccatggtgcgagccatctctcttaggcgcttgctgctgctgctgcagctgtcacaac
823 tcctagctgtcactcaagggaagacgctggtgctggggaaggaaggggaatcagcagaactgccctgcgagagttccc
824 agaagaagatcacagtcttcacctggaagttctctgaccagaggaagattctggggcagcatggcaaaggtgtattaatta
825 gaggaggttcgccttcgcagtttgatcgttttgattccaaaaaaggggcatgggagaaaggatcgtttcctctcatcatcaat
826 aaacttaagatggaagactctcagacttatatctgtgagctggagaacaggaaagaggaggtggagttgtgggtgttcaa
827 agtgaccttcagtccgggtaccagcctgttgcaagggcagagcctgaccctgaccttggatagcaactctaaggtctctaa
828 ccccttgacagagtgcaaacacaaaaagggtaaagttgtcagtggttccaaagttctctccatgtccaacctaagggttca
829 ggacagcgacttctggaactgcaccgtgaccctggaccagaaaaagaactggttcggcatgacactctcagtgctgggtt
830 ttcagagcacagctatcacggcctataagagtgagggagagtcagcggagttctccttcccactcaactttgcagaggaa
831 aacgggtggggagagctgatgtggaaggcagagaaggattctttcttccagccctggatctccttctccataaagaacaa
832 agaggtgtccgtacaaaagtccaccaaagacctcaagctccagctgaaggaaacgctcccactcaccctcaagatacc
833 ccaggtctcgcttcagtttgctggttctggcaacctgactctgactctggacaaagggacactgcatcaggaagtgaacctg
834 gtggtgatgaaagtggctcagctcaacaatactttgacctgtgaggtgatgggacctacctctcccaagatgagactgacc
835 ctgaagcaggagaaccaggaggccagggtctctgaggagcagaaagtagttcaagtggtggcccctgagacagggct
836 gtggcagtgtctactgagtgaaggtgataaggtcaagatggactccaggatccaggttttatccagaggggtgaaccaga
837 cagtgttcctggcttgcgtgctgggtggctccttcggctttctgggtttccttgggctctgcatcctctgc**TCTgtcaggTCC**c
838 ggcaccaacagcgccaggcagcacgaatgtctcagatcaagaggctcctcagtgagaagaagacctgccagtgcccc
839 caccggatgcagaagagccataatctcatctaatgaagatcttgagtcgacaagcttgctag

840

841 **CD4 Clasp:**

842 acggaattccgctcgagcgccaccatggtgcgagccatctctcttaggcgcttgctgctgctgctgcagctgtcacaac
843 tcctagctgtcactcaagggaagacgctggtgctggggaaggaaggggaatcagcagaactgccctgcgagagttccc
844 agaagaagatcacagtcttcacctggaagttctctgaccagaggaagattctggggcagcatggcaaaggtgtattaatta
845 gaggaggttcgccttcgcagtttgatcgttttgattccaaaaaaggggcatgggagaaaggatcgtttcctctcatcatcaat
846 aaacttaagatggaagactctcagacttatatctgtgagctggagaacaggaaagaggaggtggagttgtgggtgttcaa

847 agtgaccttcagtccgggtaccagcctgttgcaagggcagagcctgaccctgaccttggatagcaactctaaggtctctaa
848 ccccttgacagagtgcaaacacaaaaagggtaaagttgtcagtggttccaaagttctctccatgtccaacctaagggttca
849 ggacagcgacttctggaactgcaccgtgaccctggaccagaaaaagaactggttcggcatgacactctcagtgctgggtt
850 ttcagagcacagctatcacggcctataagagtgagggagagtcagcggagttctccttcccactcaactttgcagaggaa
851 aacgggtggggagagctgatgtggaaggcagagaaggattctttcttccagccctggatctccttctccataaagaacaa
852 agaggtgtccgtacaaaagtccaccaaagacctcaagctccagctgaaggaaacgctcccactcaccctcaagatacc
853 ccaggtctcgcttcagtttgctggttctggcaacctgactctgactctggacaaagggacactgcatcaggaagtgaacctg
854 gtggtgatgaaagtggctcagctcaacaatactttgacctgtgaggtgatgggacctacctctcccaagatgagactgacc
855 ctgaagcaggagaaccaggaggccagggtctctgaggagcagaaagtagttcaagtggtggcccctgagacagggct
856 gtggcagtgtctactgagtgaaggtgataaggtcaagatggactccaggatccaggttttatccagaggggtgaaccaga
857 cagtgttcctggcttgcgtgctgggtggctccttcggctttctgggtttccttgggctctgcatcctctgctgtgtcaggtgccggc
858 accaacagcgccaggcagcacgaatgtctcagatcaagaggctcctcagtgagaagaagacc**TCCcagTCC**ccc
859 caccggatgcagaagagccataatctcatctaatgaagatcttgagtcgacaagcttgctag
860
861 **CD4 TP**:
862 acggaattccgctcgagcgccaccatggtgcgagccatctctcttaggcgcttgctgctgctgctgctgcagctgtcacaac
863 tcctagctgtcactcaagggaagacgctggtgctggggaaggaaggggaatcagcagaactgccctgcgagagttccc
864 agaagaagatcacagtcttcacctggaagttctctgaccagaggaagattctggggcagcatggcaaaggtgtattaatta
865 gaggaggttcgccttcgcagtttgatcgttttgattccaaaaaaggggcatgggagaaaggatcgtttcctctcatcatcaat
866 aaacttaagatggaagactctcagacttatatctgtgagctggagaacaggaaagaggaggtggagttgtgggtgttcaa
867 agtgaccttcagtccgggtaccagcctgttgcaagggcagagcctgaccctgaccttggatagcaactctaaggtctctaa
868 ccccttgacagagtgcaaacacaaaaagggtaaagttgtcagtggttccaaagttctctccatgtccaacctaagggttca
869 ggacagcgacttctggaactgcaccgtgaccctggaccagaaaaagaactggttcggcatgacactctcagtgctgggtt
870 ttcagagcacagctatcacggcctataagagtgagggagagtcagcggagttctccttcccactcaactttgcagaggaa
871 aacgggtggggagagctgatgtggaaggcagagaaggattctttcttccagccctggatctccttctccataaagaacaa
872 agaggtgtccgtacaaaagtccaccaaagacctcaagctccagctgaaggaaacgctcccactcaccctcaagatacc
873 ccaggtctcgcttcagtttgctggttctggcaacctgactctgactctggacaaagggacactgcatcaggaagtgaacctg
874 gtggtgatgaaagtggctcagctcaacaatactttgacctgtgaggtgatgggacctacctctcccaagatgagactgacc
875 ctgaagcaggagaaccaggaggccagggtctctgaggagcagaaagtagttcaagtggtggcccctgagacagggct
876 gtggcagtgtctactgagtgaaggtgataaggtcaagatggactccaggatccaggttttatccagaggggtgaaccaga
877 cagtgttcctggcttgcgtgctgggt**GTGtccttcCTC**tttctgggtttccttgggctctgcatcctctgc**TCTgtcaggTC**
878 **C**cggcaccaacagcgccaggcagcacgaatgtctcagatcaagaggctcctcagtgagaagaagacctgccagtgc
879 ccccaccggatgcagaagagccataatctcatctaatgaagatcttgagtcgacaagcttgctag
880
881 **CD4 TPC**:
882 acggaattccgctcgagcgccaccatggtgcgagccatctctcttaggcgcttgctgctgctgctgctgcagctgtcacaac
883 tcctagctgtcactcaagggaagacgctggtgctggggaaggaaggggaatcagcagaactgccctgcgagagttccc
884 agaagaagatcacagtcttcacctggaagttctctgaccagaggaagattctggggcagcatggcaaaggtgtattaatta
885 gaggaggttcgccttcgcagtttgatcgttttgattccaaaaaaggggcatgggagaaaggatcgtttcctctcatcatcaat
886 aaacttaagatggaagactctcagacttatatctgtgagctggagaacaggaaagaggaggtggagttgtgggtgttcaa
887 agtgaccttcagtccgggtaccagcctgttgcaagggcagagcctgaccctgaccttggatagcaactctaaggtctctaa
888 ccccttgacagagtgcaaacacaaaaagggtaaagttgtcagtggttccaaagttctctccatgtccaacctaagggttca
889 ggacagcgacttctggaactgcaccgtgaccctggaccagaaaaagaactggttcggcatgacactctcagtgctgggtt
890 ttcagagcacagctatcacggcctataagagtgagggagagtcagcggagttctccttcccactcaactttgcagaggaa
891 aacgggtggggagagctgatgtggaaggcagagaaggattctttcttccagccctggatctccttctccataaagaacaa
892 agaggtgtccgtacaaaagtccaccaaagacctcaagctccagctgaaggaaacgctcccactcaccctcaagatacc
893 ccaggtctcgcttcagtttgctggttctggcaacctgactctgactctggacaaagggacactgcatcaggaagtgaacctg

894 gtggtgatgaaagtggctcagctcaacaatactttgacctgtgaggtgatgggacctacctctcccaagatgagactgacc
895 ctgaagcaggagaaccaggaggccagggtctctgaggagcagaaagtagttcaagtggtggcccctgagacagggct
896 gtggcagtgtctactgagtgaaggtgataaggtcaagatggactccaggatccaggttttatccagaggggtgaaccaga
897 cagtgttcctggcttgcgtgctgggt**GTGtccttcCTC**tttctgggtttccttgggctctgcatcctctgc**TCTgtcaggTC**
898 **C**cggcaccaacagcgccaggcagcacgaatgtctcagatcaagaggctcctcagtgagaagaagacc**TCCcagT**
899 **CC**ccccaccggatgcagaagagccataatctcatctaatgaagatcttgagtcgacaagcttgctag
900

901 Gene blocks were purchased from Twist Biosciences that encode the required
902 mutations in each of the following CD4 genes: R426A, R449A, K452A, Dbl, H, HC, LL,
903 SS, LL+TP, LL+R426A, and LL+C. The gene segments were cloned into pUC18
904 containing CD4 WT whereby the gene segments flanked by BamHI and HindIII were
905 replaced with the mutant geneblocks. The genes were sequence verified in pUC 18,
906 subcloned into pP2 puromycin-resistance MSCV vector (MCS-IRES-puro) via 5'XhoI
907 and 3'BglII, midi-prepped, and sequenced. The sequences for these constructs are
908 listed below. Mutated codons are shown in bold uppercase letters, while the motif under
909 interrogation is shown in bold and underlined (any non-mutated codons within the motif
910 are bold lowercase letters and underlined):

911
912
913
914 **CD4 R426A:**
915 acggaattccgctcgagcgccaccatggtgcgagccatctctcttaggcgcttgctgctgctgctgctgcagctgtcacaac
916 tcctagctgtcactcaagggaagacgctggtgctggggaaggaaggggaatcagcagaactgccctgcgagagttccc
917 agaagaagatcacagtcttcacctggaagttctctgaccagaggaagattctggggcagcatggcaaaggtgtattaatta
918 gaggaggttcgccttcgcagtttgatcgttttgattccaaaaaaggggcatgggagaaaggatcgtttcctctcatcatcaat
919 aaacttaagatggaagactctcagacttatatctgtgagctggagaacaggaaagaggaggtggagttgtgggtgttcaa
920 agtgaccttcagtccgggtaccagcctgttgcaagggcagagcctgaccctgaccttggatagcaactctaaggtctctaa
921 ccccttgacagagtgcaaacacaaaaagggtaaagttgtcagtggttccaaagttctctccatgtccaacctaagggttca
922 ggacagcgacttctggaactgcaccgtgaccctggaccagaaaaagaactggttcggcatgacactctcagtgctgggtt
923 ttcagagcacagctatcacggcctataagagtgagggagagtcagcggagttctccttcccactcaactttgcagaggaa
924 aacgggtggggagagctgatgtggaaggcagagaaggattctttcttccagccctggatctccttctccataaagaacaa
925 agaggtgtccgtacaaaagtccaccaaagacctcaagctccagctgaaggaaacgctcccactcaccctcaagatacc
926 ccaggtctcgcttcagtttgctggttctggcaacctgactctgactctggacaaagggacactgcatcaggaagtgaacctg
927 gtggtgatgaaagtggctcagctcaacaatactttgacctgtgaggtgatgggacctacctctcccaagatgagactgacc
928 ctgaagcaggagaaccaggaggccagggtctctgaggagcagaaagtagttcaagtggtggcccctgagacagggct
929 gtggcagtgtctactgagtgaaggtgataaggtcaagatggactccaggatccaggttttatccagaggggtgaaccaga
930 cagtgttcctggcttgcgtgctgggtggctccttcggctttctgggtttccttgggctctgcatcctctgctgtgtcaggtgccgg**c**
931 **accaacagGCT**caggcagcacgaatgtctcagatcaagaggctcctcagtgagaagaagacctgccagtgccccc
932 accggatgcagaagagccataatctcatctgataaagatcttgagtcgacaagcttgctag
933
934 **CD4 R449A:**
935 acggaattccgctcgagcgccaccatggtgcgagccatctctcttaggcgcttgctgctgctgctgctgcagctgtcacaac
936 tcctagctgtcactcaagggaagacgctggtgctggggaaggaaggggaatcagcagaactgccctgcgagagttccc
937 agaagaagatcacagtcttcacctggaagttctctgaccagaggaagattctggggcagcatggcaaaggtgtattaatta
938 gaggaggttcgccttcgcagtttgatcgttttgattccaaaaaaggggcatgggagaaaggatcgtttcctctcatcatcaat

939 aaacttaagatggaagactctcagacttatatctgtgagctggagaacaggaaagaggaggtggagttgtgggtgttcaa
940 agtgaccttcagtccgggtaccagcctgttgcaagggcagagcctgaccctgaccttggatagcaactctaaggtctctaa
941 ccccttgacagagtgcaaacacaaaaagggtaaagttgtcagtggttccaaagttctctccatgtccaacctaagggttca
942 ggacagcgacttctggaactgcaccgtgaccctggaccagaaaaagaactggttcggcatgacactctcagtgctgggtt
943 ttcagagcacagctatcacggcctataagagtgagggagagtcagcggagttctccttcccactcaactttgcagaggaa
944 aacgggtggggagagctgatgtggaaggcagagaaggattctttcttccagccctggatctccttctccataaagaacaa
945 agaggtgtccgtacaaaagtccaccaaagacctcaagctccagctgaaggaaacgctcccactcaccctcaagatacc
946 ccaggtctcgcttcagtttgctggttctggcaacctgactctgactctggacaaagggacactgcatcaggaagtgaacctg
947 gtggtgatgaaagtggctcagctcaacaatactttgacctgtgaggtgatgggacctacctctcccaagatgagactgacc
948 ctgaagcaggagaaccaggaggccagggtctctgaggagcagaaagtagttcaagtggtggcccctgagacagggct
949 gtggcagtgtctactgagtgaaggtgataaggtcaagatggactccaggatccaggttttatccagaggggtgaaccaga
950 cagtgttcctggcttgcgtgctgggtggctccttcggctttctgggtttccttgggctctgcatcctctgctgtgtcaggtgccggc
951 accaacagcgccaggcagcacgaatgtctcagatcaagaggctcctcagtgagaagaagacctgccagtgcccc**cac**
952 **GCTatgcagaag**agccataatctcatctgataaagatcttgagtcgacaagcttgctag
953
954 **CD4 K452A**:
955
956 acggaattccgctcgagcgccaccatggtgcgagccatctctcttaggcgcttgctgctgctgctgctgcagctgtcacaac
957 tcctagctgtcactcaagggaagacgctggtgctggggaaggaaggggaatcagcagaactgccctgcgagagttccc
958 agaagaagatcacagtcttcacctggaagttctctgaccagaggaagattctggggcagcatggcaaaggtgtattaatta
959 gaggaggttcgccttcgcagtttgatcgttttgattccaaaaaaggggcatgggagaaaggatcgtttcctctcatcatcaat
960 aaacttaagatggaagactctcagacttatatctgtgagctggagaacaggaaagaggaggtggagttgtgggtgttcaa
961 agtgaccttcagtccgggtaccagcctgttgcaagggcagagcctgaccctgaccttggatagcaactctaaggtctctaa
962 ccccttgacagagtgcaaacacaaaaagggtaaagttgtcagtggttccaaagttctctccatgtccaacctaagggttca
963 ggacagcgacttctggaactgcaccgtgaccctggaccagaaaaagaactggttcggcatgacactctcagtgctgggtt
964 ttcagagcacagctatcacggcctataagagtgagggagagtcagcggagttctccttcccactcaactttgcagaggaa
965 aacgggtggggagagctgatgtggaaggcagagaaggattctttcttccagccctggatctccttctccataaagaacaa
966 agaggtgtccgtacaaaagtccaccaaagacctcaagctccagctgaaggaaacgctcccactcaccctcaagatacc
967 ccaggtctcgcttcagtttgctggttctggcaacctgactctgactctggacaaagggacactgcatcaggaagtgaacctg
968 gtggtgatgaaagtggctcagctcaacaatactttgacctgtgaggtgatgggacctacctctcccaagatgagactgacc
969 ctgaagcaggagaaccaggaggccagggtctctgaggagcagaaagtagttcaagtggtggcccctgagacagggct
970 gtggcagtgtctactgagtgaaggtgataaggtcaagatggactccaggatccaggttttatccagaggggtgaaccaga
971 cagtgttcctggcttgcgtgctgggtggctccttcggctttctgggtttccttgggctctgcatcctctgctgtgtcaggtgccggc
972 accaacagcgccaggcagcacgaatgtctcagatcaagaggctcctcagtgagaagaagacctgccagtgcccc**cac**
973 **cggatgcagGCA**agccataatctcatctgataaagatcttgagtcgacaagcttgctag
974
975 **CD4 Dbl**:
976 acggaattccgctcgagcgccaccatggtgcgagccatctctcttaggcgcttgctgctgctgctgctgcagctgtcacaac
977 tcctagctgtcactcaagggaagacgctggtgctggggaaggaaggggaatcagcagaactgccctgcgagagttccc
978 agaagaagatcacagtcttcacctggaagttctctgaccagaggaagattctggggcagcatggcaaaggtgtattaatta
979 gaggaggttcgccttcgcagtttgatcgttttgattccaaaaaaggggcatgggagaaaggatcgtttcctctcatcatcaat
980 aaacttaagatggaagactctcagacttatatctgtgagctggagaacaggaaagaggaggtggagttgtgggtgttcaa
981 agtgaccttcagtccgggtaccagcctgttgcaagggcagagcctgaccctgaccttggatagcaactctaaggtctctaa
982 ccccttgacagagtgcaaacacaaaaagggtaaagttgtcagtggttccaaagttctctccatgtccaacctaagggttca
983 ggacagcgacttctggaactgcaccgtgaccctggaccagaaaaagaactggttcggcatgacactctcagtgctgggtt
984 ttcagagcacagctatcacggcctataagagtgagggagagtcagcggagttctccttcccactcaactttgcagaggaa
985 aacgggtggggagagctgatgtggaaggcagagaaggattctttcttccagccctggatctccttctccataaagaacaa

986 agaggtgtccgtacaaaagtccaccaaagacctcaagctccagctgaaggaaacgctcccactcaccctcaagatacc
987 ccaggtctcgcttcagtttgctggttctggcaacctgactctgactctggacaaagggacactgcatcaggaagtgaacctg
988 gtggtgatgaaagtggctcagctcaacaatactttgacctgtgaggtgatgggacctacctctcccaagatgagactgacc
989 ctgaagcaggagaaccaggaggccagggtctctgaggagcagaaagtagttcaagtggtggcccctgagacagggct
990 gtggcagtgtctactgagtgaaggtgataaggtcaagatggactccaggatccaggttttatccagaggggtgaaccaga
991 cagtgttcctggcttgcgtgctgggtggctccttcggctttctgggtttccttgggctctgcatcctctgctgtgtcaggtgccggc
992 accaacagcgccaggcagcacgaatgtctcagatcaagaggctcctcagtgagaagaagacctgccagtgcccc**cac**
993 **GCTatgcagGCA**agccataatctcatctgataaagatcttgagtcgacaagcttgctag
994
995 **CD4 H**:
996 acggaattccgctcgagcgccaccatggtgcgagccatctctcttaggcgcttgctgctgctgctgctgcagctgtcacaac
997 tcctagctgtcactcaagggaagacgctggtgctggggaaggaaggggaatcagcagaactgccctgcgagagttccc
998 agaagaagatcacagtcttcacctggaagttctctgaccagaggaagattctggggcagcatggcaaaggtgtattaatta
999 gaggaggttcgccttcgcagtttgatcgttttgattccaaaaaaggggcatgggagaaaggatcgtttcctctcatcatcaat
1000 aaacttaagatggaagactctcagacttatatctgtgagctggagaacaggaaagaggaggtggagttgtgggtgttcaa
1001 agtgaccttcagtccgggtaccagcctgttgcaagggcagagcctgaccctgaccttggatagcaactctaaggtctctaa
1002 ccccttgacagagtgcaaacacaaaaagggtaaagttgtcagtggttccaaagttctctccatgtccaacctaagggttca
1003 ggacagcgacttctggaactgcaccgtgaccctggaccagaaaaagaactggttcggcatgacactctcagtgctgggtt
1004 ttcagagcacagctatcacggcctataagagtgagggagagtcagcggagttctccttcccactcaactttgcagaggaa
1005 aacgggtggggagagctgatgtggaaggcagagaaggattctttcttccagccctggatctccttctccataaagaacaa
1006 agaggtgtccgtacaaaagtccaccaaagacctcaagctccagctgaaggaaacgctcccactcaccctcaagatacc
1007 ccaggtctcgcttcagtttgctggttctggcaacctgactctgactctggacaaagggacactgcatcaggaagtgaacctg
1008 gtggtgatgaaagtggctcagctcaacaatactttgacctgtgaggtgatgggacctacctctcccaagatgagactgacc
1009 ctgaagcaggagaaccaggaggccagggtctctgaggagcagaaagtagttcaagtggtggcccctgagacagggct
1010 gtggcagtgtctactgagtgaaggtgataaggtcaagatggactccaggatccaggttttatccagaggggtgaaccaga
1011 cagtgttcctggcttgcgtgctgggtggctccttcggctttctgggtttccttgggctctgcatcctctgctgtgtcaggtgccggc
1012 accaacagcgccaggcagca**AACGGCCCTGGTGGCAATCCTGGTGGCAACGCAGGCGG**
1013 **T**acctgccagtgcccccaccggatgcagaagagccataatctcatctgataaagatcttgagtcgacaagcttgctag
1014
1015 **CD4 HC**:
1016 acggaattccgctcgagcgccaccatggtgcgagccatctctcttaggcgcttgctgctgctgctgctgcagctgtcacaac
1017 tcctagctgtcactcaagggaagacgctggtgctggggaaggaaggggaatcagcagaactgccctgcgagagttccc
1018 agaagaagatcacagtcttcacctggaagttctctgaccagaggaagattctggggcagcatggcaaaggtgtattaatta
1019 gaggaggttcgccttcgcagtttgatcgttttgattccaaaaaaggggcatgggagaaaggatcgtttcctctcatcatcaat
1020 aaacttaagatggaagactctcagacttatatctgtgagctggagaacaggaaagaggaggtggagttgtgggtgttcaa
1021 agtgaccttcagtccgggtaccagcctgttgcaagggcagagcctgaccctgaccttggatagcaactctaaggtctctaa
1022 ccccttgacagagtgcaaacacaaaaagggtaaagttgtcagtggttccaaagttctctccatgtccaacctaagggttca
1023 ggacagcgacttctggaactgcaccgtgaccctggaccagaaaaagaactggttcggcatgacactctcagtgctgggtt
1024 ttcagagcacagctatcacggcctataagagtgagggagagtcagcggagttctccttcccactcaactttgcagaggaa
1025 aacgggtggggagagctgatgtggaaggcagagaaggattctttcttccagccctggatctccttctccataaagaacaa
1026 agaggtgtccgtacaaaagtccaccaaagacctcaagctccagctgaaggaaacgctcccactcaccctcaagatacc
1027 ccaggtctcgcttcagtttgctggttctggcaacctgactctgactctggacaaagggacactgcatcaggaagtgaacctg
1028 gtggtgatgaaagtggctcagctcaacaatactttgacctgtgaggtgatgggacctacctctcccaagatgagactgacc
1029 ctgaagcaggagaaccaggaggccagggtctctgaggagcagaaagtagttcaagtggtggcccctgagacagggct
1030 gtggcagtgtctactgagtgaaggtgataaggtcaagatggactccaggatccaggttttatccagaggggtgaaccaga
1031 cagtgttcctggcttgcgtgctgggtggctccttcggctttctgggtttccttgggctctgcatcctctgctgtgtcaggtgccggc
1032 accaacagcgccaggcagca**AACGGCCCTGGTGGCAATCCTGGTGGCAACGCAGGCGG**

1033 **T**acc**TCCcagTCC**ccccaccggatgcagaagagccataatctcatctgataaagatcttgagtcgacaagcttgcta
1034 g
1035
1036 **CD4 LL**:
1037 acggaattccgctcgagcgccaccatggtgcgagccatctctcttaggcgcttgctgctgctgctgctgcagctgtcacaac
1038 tcctagctgtcactcaagggaagacgctggtgctggggaaggaaggggaatcagcagaactgccctgcgagagttccc
1039 agaagaagatcacagtcttcacctggaagttctctgaccagaggaagattctggggcagcatggcaaaggtgtattaatta
1040 gaggaggttcgccttcgcagtttgatcgttttgattccaaaaaaggggcatgggagaaaggatcgtttcctctcatcatcaat
1041 aaacttaagatggaagactctcagacttatatctgtgagctggagaacaggaaagaggaggtggagttgtgggtgttcaa
1042 agtgaccttcagtccgggtaccagcctgttgcaagggcagagcctgaccctgaccttggatagcaactctaaggtctctaa
1043 ccccttgacagagtgcaaacacaaaaagggtaaagttgtcagtggttccaaagttctctccatgtccaacctaagggttca
1044 ggacagcgacttctggaactgcaccgtgaccctggaccagaaaaagaactggttcggcatgacactctcagtgctgggtt
1045 ttcagagcacagctatcacggcctataagagtgagggagagtcagcggagttctccttcccactcaactttgcagaggaa
1046 aacgggtggggagagctgatgtggaaggcagagaaggattctttcttccagccctggatctccttctccataaagaacaa
1047 agaggtgtccgtacaaaagtccaccaaagacctcaagctccagctgaaggaaacgctcccactcaccctcaagatacc
1048 ccaggtctcgcttcagtttgctggttctggcaacctgactctgactctggacaaagggacactgcatcaggaagtgaacctg
1049 gtggtgatgaaagtggctcagctcaacaatactttgacctgtgaggtgatgggacctacctctcccaagatgagactgacc
1050 ctgaagcaggagaaccaggaggccagggtctctgaggagcagaaagtagttcaagtggtggcccctgagacagggct
1051 gtggcagtgtctactgagtgaaggtgataaggtcaagatggactccaggatccaggttttatccagaggggtgaaccaga
1052 cagtgttcctggcttgcgtgctgggtggctccttcggctttctgggtttccttgggctctgcatcctctgctgtgtcaggtgccggc
1053 accaacagcgccaggcagcacgaatgtctcag**atcaagaggGCTGCA**agtgagaagaagacctgccagtgccc
1054 ccaccggatgcagaagagccataatctcatctgataaagatcttgagtcgacaagcttgctag
1055
1056 **CD4 SS**:
1057 acggaattccgctcgagcgccaccatggtgcgagccatctctcttaggcgcttgctgctgctgctgctgcagctgtcacaac
1058 tcctagctgtcactcaagggaagacgctggtgctggggaaggaaggggaatcagcagaactgccctgcgagagttccc
1059 agaagaagatcacagtcttcacctggaagttctctgaccagaggaagattctggggcagcatggcaaaggtgtattaatta
1060 gaggaggttcgccttcgcagtttgatcgttttgattccaaaaaaggggcatgggagaaaggatcgtttcctctcatcatcaat
1061 aaacttaagatggaagactctcagacttatatctgtgagctggagaacaggaaagaggaggtggagttgtgggtgttcaa
1062 agtgaccttcagtccgggtaccagcctgttgcaagggcagagcctgaccctgaccttggatagcaactctaaggtctctaa
1063 ccccttgacagagtgcaaacacaaaaagggtaaagttgtcagtggttccaaagttctctccatgtccaacctaagggttca
1064 ggacagcgacttctggaactgcaccgtgaccctggaccagaaaaagaactggttcggcatgacactctcagtgctgggtt
1065 ttcagagcacagctatcacggcctataagagtgagggagagtcagcggagttctccttcccactcaactttgcagaggaa
1066 aacgggtggggagagctgatgtggaaggcagagaaggattctttcttccagccctggatctccttctccataaagaacaa
1067 agaggtgtccgtacaaaagtccaccaaagacctcaagctccagctgaaggaaacgctcccactcaccctcaagatacc
1068 ccaggtctcgcttcagtttgctggttctggcaacctgactctgactctggacaaagggacactgcatcaggaagtgaacctg
1069 gtggtgatgaaagtggctcagctcaacaatactttgacctgtgaggtgatgggacctacctctcccaagatgagactgacc
1070 ctgaagcaggagaaccaggaggccagggtctctgaggagcagaaagtagttcaagtggtggcccctgagacagggct
1071 gtggcagtgtctactgagtgaaggtgataaggtcaagatggactccaggatccaggttttatccagaggggtgaaccaga
1072 cagtgttcctggcttgcgtgctgggtggctccttcggctttctgggtttccttgggctctgcatcctctgctgtgtcaggtgccggc
1073 accaacagcgccaggcagcacgaatg**GCTcagatcaagaggctcctcGCA**gagaagaagacctgccagtgcc
1074 cccaccggatgcagaagagccataatctcatctgataaagatcttgagtcgacaagcttgctag
1075
1076 **CD4LL+TP**:
1077 acggaattccgctcgagcgccaccatggtgcgagccatctctcttaggcgcttgctgctgctgctgctgcagctgtcacaac
1078 tcctagctgtcactcaagggaagacgctggtgctggggaaggaaggggaatcagcagaactgccctgcgagagttccc
1079 agaagaagatcacagtcttcacctggaagttctctgaccagaggaagattctggggcagcatggcaaaggtgtattaatta

1080 gaggaggttcgccttcgcagtttgatcgttttgattccaaaaaaggggcatgggagaaaggatcgtttcctctcatcatcaat
1081 aaacttaagatggaagactctcagacttatatctgtgagctggagaacaggaaagaggaggtggagttgtgggtgttcaa
1082 agtgaccttcagtccgggtaccagcctgttgcaagggcagagcctgaccctgaccttggatagcaactctaaggtctctaa
1083 ccccttgacagagtgcaaacacaaaaagggtaaagttgtcagtggttccaaagttctctccatgtccaacctaagggttca
1084 ggacagcgacttctggaactgcaccgtgaccctggaccagaaaaagaactggttcggcatgacactctcagtgctgggtt
1085 ttcagagcacagctatcacggcctataagagtgagggagagtcagcggagttctccttcccactcaactttgcagaggaa
1086 aacgggtggggagagctgatgtggaaggcagagaaggattctttcttccagccctggatctccttctccataaagaacaa
1087 agaggtgtccgtacaaaagtccaccaaagacctcaagctccagctgaaggaaacgctcccactcaccctcaagatacc
1088 ccaggtctcgcttcagtttgctggttctggcaacctgactctgactctggacaaagggacactgcatcaggaagtgaacctg
1089 gtggtgatgaaagtggctcagctcaacaatactttgacctgtgaggtgatgggacctacctctcccaagatgagactgacc
1090 ctgaagcaggagaaccaggaggccagggtctctgaggagcagaaagtagttcaagtggtggcccctgagacagggct
1091 gtggcagtgtctactgagtgaaggtgataaggtcaagatggactccaggatccaggttttatccagaggggtgaaccaga
1092 cagtgttcctggcttgcgtgctgggt**GTGtccttcCTC**tttctgggtttccttgggctctgcatcctctgc**TCTgtcaggTC**
1093 **C**cggcaccaacagcgccaggcagcacgaatgtctcag**atcaagaggGCTGCA**agtgagaagaagacctgcca
1094 gtgcccccaccggatgcagaagagccataatctcatctaatgaagatcttgagtcgacaagcttgctag
1095
1096 **CD4LL+R426A**:
1097 acggaattccgctcgagcgccaccatggtgcgagccatctctcttaggcgcttgctgctgctgctgctgcagctgtcacaac
1098 tcctagctgtcactcaagggaagacgctggtgctggggaaggaaggggaatcagcagaactgccctgcgagagttccc
1099 agaagaagatcacagtcttcacctggaagttctctgaccagaggaagattctggggcagcatggcaaaggtgtattaatta
1100 gaggaggttcgccttcgcagtttgatcgttttgattccaaaaaaggggcatgggagaaaggatcgtttcctctcatcatcaat
1101 aaacttaagatggaagactctcagacttatatctgtgagctggagaacaggaaagaggaggtggagttgtgggtgttcaa
1102 agtgaccttcagtccgggtaccagcctgttgcaagggcagagcctgaccctgaccttggatagcaactctaaggtctctaa
1103 ccccttgacagagtgcaaacacaaaaagggtaaagttgtcagtggttccaaagttctctccatgtccaacctaagggttca
1104 ggacagcgacttctggaactgcaccgtgaccctggaccagaaaaagaactggttcggcatgacactctcagtgctgggtt
1105 ttcagagcacagctatcacggcctataagagtgagggagagtcagcggagttctccttcccactcaactttgcagaggaa
1106 aacgggtggggagagctgatgtggaaggcagagaaggattctttcttccagccctggatctccttctccataaagaacaa
1107 agaggtgtccgtacaaaagtccaccaaagacctcaagctccagctgaaggaaacgctcccactcaccctcaagatacc
1108 ccaggtctcgcttcagtttgctggttctggcaacctgactctgactctggacaaagggacactgcatcaggaagtgaacctg
1109 gtggtgatgaaagtggctcagctcaacaatactttgacctgtgaggtgatgggacctacctctcccaagatgagactgacc
1110 ctgaagcaggagaaccaggaggccagggtctctgaggagcagaaagtagttcaagtggtggcccctgagacagggct
1111 gtggcagtgtctactgagtgaaggtgataaggtcaagatggactccaggatccaggttttatccagaggggtgaaccaga
1112 cagtgttcctggcttgcgtgctgggtggctccttcggctttctgggtttccttgggctctgcatcctctgctgtgtcaggtgccgg**c**
1113 **accaacagGCT**caggcagcacgaatgtctcag**atcaagaggGCTGCA**agtgagaagaagacctgccagtgc
1114 ccccaccggatgcagaagagccataatctcatctgataaagatcttgagtcgacaagcttgctag
1115
1116 **CD4LL+C**:
1117 acggaattccgctcgagcgccaccatggtgcgagccatctctcttaggcgcttgctgctgctgctgctgcagctgtcacaac
1118 tcctagctgtcactcaagggaagacgctggtgctggggaaggaaggggaatcagcagaactgccctgcgagagttccc
1119 agaagaagatcacagtcttcacctggaagttctctgaccagaggaagattctggggcagcatggcaaaggtgtattaatta
1120 gaggaggttcgccttcgcagtttgatcgttttgattccaaaaaaggggcatgggagaaaggatcgtttcctctcatcatcaat
1121 aaacttaagatggaagactctcagacttatatctgtgagctggagaacaggaaagaggaggtggagttgtgggtgttcaa
1122 agtgaccttcagtccgggtaccagcctgttgcaagggcagagcctgaccctgaccttggatagcaactctaaggtctctaa
1123 ccccttgacagagtgcaaacacaaaaagggtaaagttgtcagtggttccaaagttctctccatgtccaacctaagggttca
1124 ggacagcgacttctggaactgcaccgtgaccctggaccagaaaaagaactggttcggcatgacactctcagtgctgggtt
1125 ttcagagcacagctatcacggcctataagagtgagggagagtcagcggagttctccttcccactcaactttgcagaggaa

aacgggtggggagagctgatgtggaaggcagagaaggattctttcttccagccctggatctccttctccataaagaacaa
agaggtgtccgtacaaaagtccaccaaagacctcaagctccagctgaaggaaacgctcccactcaccctcaagatacc
ccaggtctcgcttcagtttgctggttctggcaacctgactctgactctggacaaagggacactgcatcaggaagtgaacctg
gtggtgatgaaagtggctcagctcaacaatactttgacctgtgaggtgatgggacctacctctcccaagatgagactgacc
ctgaagcaggagaaccaggaggccagggtctctgaggagcagaaagtagttcaagtggtggcccctgagacagggct
gtggcagtgtctactgagtgaaggtgataaggtcaagatggactccaggatccaggttttatccagaggggtgaaccaga
cagtgttcctggcttgcgtgctgggtggctccttcggctttctgggtttccttgggctctgcatcctctgctgtgtcaggtgccggc
accaacagcgccaggcagcacgaatgtctcag**atcaagaggGCTGCA**agtgagaagaagacc**TCCcagTC**
Cccccaccggatgcagaagagccataatctcatctgataaagatcttgagtcgacaagcttgctag

The genes of CD4LL+PLXF and CD4LL+Δbind were made by cloning the in frame extraceullar domain of CD4 that encodes either the PLXF to ELXF (Glassman et al., 2018) or Δbind extracellular domain (Parrish et al., 2015) flanked by XhoI and BamHI into the CD4LL gene. The genes were sequence verified in pUC 18, subcloned into pP2 puromycin-resistance MSCV vector (MCS-IRES-puro) via 5'XhoI and 3'BglII, midi-prepped, and sequenced. The sequences for these constructs are as follows with mutated codons shown in bold uppercase letters, while the motif under interrogation is shown in bold and underlined (any non-mutated codons within the motif are bold lowercase letters and underlined):

CD4LL+PLXF:
acggaattccgctcgagcgccaccatggtgcgagccatctctcttaggcgcttgctgctgctgctgctgcagctgtcacaac
tcctagctgtcactcaagggaagacgctggtgctggggaaggaaggggaatcagcagaactgccctgcgagagttccc
agaagaagatcacagtcttcacctggaagttctctgaccagaggaagattctggggcagcatggcaaaggtgtattaatta
gaggaggttcgccttcgcagtttgatcgttttgattccaaaaaaggggcatgggagaaaggatcgtttcctctcatcatcaat
aaacttaagatggaagactctcagacttatatctgtgagctggagaacaggaaagaggaggtggagttgtgggtgttcaa
agtgaccttcagtccgggtaccagcctgttgcaagggcagagcctgaccctgaccttggatagcaactctaaggtctctaa
ccccttgacagagtgcaaacacaaaaagggtaaagttgtcagtggttccaaagttctctccatgtccaacctaagggttca
ggacagcgacttctggaactgcaccgtgaccctggaccagaaaaagaactggttcggcatgacactctcagtgctgggtt
ttcagagcacagctatcacggcctataagagtgagggagagtcagcggagttctccttc**GAActcaacGAG**gcaga
ggaaaacgggtggggagagctgatgtggaaggcagagaaggattctttcttccagccctggatctccttctccataaaga
acaaagaggtgtccgtacaaaagtccaccaaagacctcaagctccagctgaaggaaacgctcccactcaccctcaag
ataccccaggtctcgcttcagtttgctggttctggcaacctgactctgactctggacaaagggacactgcatcaggaagtga
acctggtggtgatgaaagtggctcagctcaacaatactttgacctgtgaggtgatgggacctacctctcccaagatgagact
gaccctgaagcaggagaaccaggaggccagggtctctgaggagcagaaagtagttcaagtggtggcccctgagaca
gggctgtggcagtgtctactgagtgaaggtgataaggtcaagatggactccaggatccaggttttatccagaggggtgaa
ccagacagtgttcctggcttgcgtgctgggtggctccttcggctttctgggtttccttgggctctgcatcctctgctgtgtcaggtg
ccggcaccaacagcgccaggcagcacgaatgtctcag**atcaagaggGCTGCA**agtgagaagaagacctgcca
gtgcccccaccggatgcagaagagccataatctcatctgataaagatcttgagtcgacaagcttgctag

CD4LL+Δbind:
acggaattccgctcgagcgccaccatggtgcgagccatctctcttaggcgcttgctgctgctgctgctgcagctgtcacaac
tcctagctgtcactcaagggaagacgctggtgctggggaaggaaggggaatcagcagaactgccctgcgagagttccc
agaagaagatcacagtcttcacctggaagttctctgaccagaggaagattctggggcagcatggc**GATggtGATTC**

1170 **AGATAGC**ggaggttcgccttcgcagtttgatcgttttgattccaaaaaaggggcatgggagaaaggatcgtttcctctc
1171 atcatcaataaacttaagatggaagactctcagacttatatctgtgagctggagaacaggaaagaggaggtggagttgtg
1172 ggtgttcaaagtgaccttcagtccgggtaccagcctgttgcaagggcagagcctgaccctgaccttggatagcaactctaa
1173 ggtctctaaccccttgacagagtgcaaacacaaaaagggtaaagttgtcagtggttccaaagttctctccatgtccaaccta
1174 agggttcaggacagcgacttctggaactgcaccgtgaccctggaccagaaaaagaactggttcggcatgacactctcag
1175 tgctgggttttcagagcacagctatcacggcctataagagtgagggagagtcagcggagttctccttcccactcaactttgc
1176 agaggaaaacgggtggggagagctgatgtggaaggcagagaaggattctttcttccagccctggatctccttctccataa
1177 agaacaaagaggtgtccgtacaaaagtccaccaaagacctcaagctccagctgaaggaaacgctcccactcaccctc
1178 aagataccccaggtctcgcttcagtttgctggttctggcaacctgactctgactctggacaaagggacactgcatcaggaa
1179 gtgaacctggtggtgatgaaagtggctcagctcaacaatactttgacctgtgaggtgatgggacctacctctcccaagatg
1180 agactgaccctgaagcaggagaaccaggaggccagggtctctgaggagcagaaagtagttcaagtggtggcccctga
1181 gacagggctgtggcagtgtctactgagtgaaggtgataaggtcaagatggactccaggatccaggttttatccagagggg
1182 tgaaccagacagtgttcctggcttgcgtgctgggtggctccttcggctttctgggtttccttgggctctgcatcctctgctgtgtca
1183 ggtgccggcaccaacagcgccaggcagcacgaatgtctcag**atcaagaggGCTGCA**agtgagaagaagacctg
1184 ccagtgcccccaccggatgcagaagagccataatctcatctgataaagatcttgagtcgacaagcttgctag
1185

1186 **REFERENCES**

1187

1188 Aivazian, D., and Stern, L.J. (2000). Phosphorylation of T cell receptor zeta is regulated by a lipid
1189 dependent folding transition. Nat Struct Biol *7*, 1023-1026.

1190 Andersen, K.G., Nissen, J.K., and Betz, A.G. (2012). Comparative Genomics Reveals Key Gain-of-
1191 Function Events in Foxp3 during Regulatory T Cell Evolution. Front Immunol *3*, 113.

1192 Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., and Ben-Tal, N. (2016).
1193 ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in
1194 macromolecules. Nucleic Acids Res *44*, W344-350.

1195 Ashkenazy, H., Erez, E., Martz, E., Pupko, T., and Ben-Tal, N. (2010). ConSurf 2010: calculating
1196 evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic
1197 Acids Res *38*, W529-533.

1198 Bernstein, R.M., Schluter, S.F., Bernstein, H., and Marchalonis, J.J. (1996). Primordial emergence
1199 of the recombination activating gene 1 (RAG1): sequence of the complete shark gene indicates
1200 homology to microbial integrases. Proc Natl Acad Sci U S A *93*, 9454-9459.

1201 Berry, D.M., Nash, P., Liu, S.K., Pawson, T., and McGlade, C.J. (2002). A high-affinity Arg-X-X-Lys
1202 SH3 binding motif confers specificity for the interaction between Gads and SLP-76 in T cell
1203 signaling. Curr Biol *12*, 1336-1341.

1204 Buslje, C.M., Santos, J., Delfino, J.M., and Nielsen, M. (2009). Correction for phylogeny, small
1205 number of observations and data redundancy improves the identification of coevolving amino
1206 acid pairs using mutual information. Bioinformatics *25*, 1125-1131.

1207 Capra, J.A., and Singh, M. (2007). Predicting functionally important residues from sequence
1208 conservation. Bioinformatics *23*, 1875-1882.

1209 Colell, E.A., Iserte, J.A., Simonetti, F.L., and Marino-Buslje, C. (2018). MISTIC2: comprehensive
1210 server to study coevolution in protein families. Nucleic Acids Res *46*, W323-W328.

1211 Crise, B., and Rose, J.K. (1992). Identification of palmitoylation sites on CD4, the human
1212 immunodeficiency virus receptor. J Biol Chem *267*, 13593-13597.

1213 Daugherty, M.D., and Malik, H.S. (2012). Rules of engagement: molecular insights from host-
1214 virus arms races. Annu Rev Genet *46*, 677-700.

1215 Davey, N.E., Cyert, M.S., and Moses, A.M. (2015). Short linear motifs - ex nihilo evolution of
1216 protein regulation. Cell Commun Signal *13*, 43.

1217 Elde, N.C., Child, S.J., Geballe, A.P., and Malik, H.S. (2009). Protein kinase R reveals an
1218 evolutionary model for defeating viral mimicry. Nature *457*, 485-489.

1219 Eshhar, Z., Waks, T., Gross, G., and Schindler, D.G. (1993). Specific activation and targeting of
1220 cytotoxic lymphocytes through chimeric single chains consisting of antibody-binding domains
1221 and the gamma or zeta subunits of the immunoglobulin and T-cell receptors. Proc Natl Acad Sci
1222 U S A *90*, 720-724.

1223 Fessler, M.B. (2016). The Intracellular Cholesterol Landscape: Dynamic Integrator of the
1224 Immune Response. Trends Immunol *37*, 819-830.

1225 Foley, G., Mora, A., Ross, C.M., Bottoms, S., Sützl, L., Lamprecht, M.L., Zaugg, J., Essebier, A.,
1226 Balderson, B., Newell, R.*, et al.* (2020). Identifying and engineering ancient variants of enzymes
1227 using Graphical Representation of Ancestral Sequence Predictions (GRASP). bioRxiv,
1228 2019.2012.2030.891457.

1229 Fragoso, R., Ren, D., Zhang, X., Su, M.W., Burakoff, S.J., and Jin, Y.J. (2003). Lipid raft distribution
1230 of CD4 depends on its palmitoylation and association with Lck, and evidence for CD4-induced
1231 lipid raft aggregation as an additional mechanism to enhance CD3 signaling. J Immunol *170*,
1232 913-921.

1233 Garron, M.L., Arthos, J., Guichou, J.F., McNally, J., Cicala, C., and Arold, S.T. (2008). Structural
1234 basis for the interaction between focal adhesion kinase and CD4. J Mol Biol *375*, 1320-1328.

1235 Glaichenhaus, N., Shastri, N., Littman, D.R., and Turner, J.M. (1991). Requirement for
1236 association of p56lck with CD4 in antigen-specific signal transduction in T cells. Cell *64*, 511-520.

1237 Glassman, C.R., Parrish, H.L., Deshpande, N.R., and Kuhns, M.S. (2016). The CD4 and
1238 CD3deltaepsilon Cytosolic Juxtamembrane Regions Are Proximal within a Compact TCR-CD3-
1239 pMHC-CD4 Macrocomplex. J Immunol *196*, 4713-4722.

1240 Glassman, C.R., Parrish, H.L., Lee, M.S., and Kuhns, M.S. (2018). Reciprocal TCR-CD3 and CD4
1241 Engagement of a Nucleating pMHCII Stabilizes a Functional Receptor Macrocomplex. Cell Rep
1242 *22*, 1263-1275.

1243 Guy, C.S., and Vignali, D.A. (2009). Organization of proximal signal initiation at the TCR:CD3
1244 complex. Immunological Reviews *232*, 7-21.

1245 Hochberg, G.K.A., and Thornton, J.W. (2017). Reconstructing Ancient Proteins to Understand
1246 the Causes of Structure and Function. Annu Rev Biophys *46*, 247-269.

1247 Holst, J., Wang, H., Eder, K.D., Workman, C.J., Boyd, K.L., Baquet, Z., Singh, H., Forbes, K.,
1248 Chruscinski, A., Smeyne, R.*, et al.* (2008). Scalable signaling mediated by T cell antigen receptor-
1249 CD3 ITAMs ensures effective negative selection and prevents autoimmunity. Nat Immunol *9*,
1250 658-666.

1251 Hur, E.M., Son, M., Lee, O.H., Choi, Y.B., Park, C., Lee, H., and Yun, Y. (2003). LIME, a novel
1252 transmembrane adaptor protein, associates with p56lck and mediates T cell activation. J Exp
1253 Med *198*, 1463-1473.

1254 Jones, D.T., Taylor, W.R., and Thornton, J.M. (1992). The rapid generation of mutation data
1255 matrices from protein sequences. Comput Appl Biosci *8*, 275-282.

1256 Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid
1257 multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res *30*, 3059-3066.

1258 Kelly, B.T., McCoy, A.J., Spate, K., Miller, S.E., Evans, P.R., Honing, S., and Owen, D.J. (2008). A
1259 structural explanation for the binding of endocytic dileucine motifs by the AP2 complex. Nature
1260 *456*, 976-979.

1261 Kim, P.W., Sun, Z.Y., Blacklow, S.C., Wagner, G., and Eck, M.J. (2003). A zinc clasp structure
1262 tethers Lck to T cell coreceptors CD4 and CD8. Science *301*, 1725-1728.

1263 Kobayashi, S., Thelin, M.A., Parrish, H.L., Deshpande, N.R., Lee, M.S., Karimzadeh, A., Niewczas,
1264 M.A., Serwold, T., and Kuhns, M.S. (2020). A biomimetic five-module chimeric antigen receptor
1265 ((5M)CAR) designed to target and eliminate antigen-specific T cells. Proc Natl Acad Sci U S A
1266 *117*, 28950-28959.

1267 Kosakovsky Pond, S.L., and Frost, S.D. (2005). Not so different after all: a comparison of
1268 methods for detecting amino acid sites under selection. Mol Biol Evol *22*, 1208-1222.

1269 Kosakovsky Pond, S.L., Poon, A.F.Y., Velazquez, R., Weaver, S., Hepler, N.L., Murrell, B., Shank,
1270 S.D., Magalis, B.R., Bouvier, D., Nekrutenko, A.*, et al.* (2020). HyPhy 2.5-A Customizable
1271 Platform for Evolutionary Hypothesis Testing Using Phylogenies. Mol Biol Evol *37*, 295-299.

1272 Kowarsch, A., Fuchs, A., Frishman, D., and Pagel, P. (2010). Correlated mutations: a hallmark of
1273 phenotypic amino acid substitutions. PLoS Comput Biol *6*.

1274 Kuhns, M.S., and Davis, M.M. (2007). Disruption of extracellular interactions impairs T cell
1275 receptor-CD3 complex stability and signaling. Immunity *26*, 357-369.

1276 Kuhns, M.S., and Davis, M.M. (2012). TCR Signaling Emerges from the Sum of Many Parts. Front
1277 Immunol *3*, 159.

1278 Kwon, Y., Kaake, R.M., Echeverria, I., Suarez, M., Karimian Shamsabadi, M., Stoneham, C.,
1279 Ramirez, P.W., Kress, J., Singh, R., Sali, A.*, et al.* (2020). Structural basis of CD4 downregulation
1280 by HIV-1 Nef. Nat Struct Mol Biol *27*, 822-828.

1281 Labanieh, L., Majzner, R.G., and Mackall, C.L. (2018). Programming CAR-T cells to kill cancer. Nat
1282 Biomed Eng *2*, 377-391.

1283 Ladygina, N., Martin, B.R., and Altman, A. (2011). Dynamic palmitoylation and the role of DHHC
1284 proteins in T cell activation and anergy. Adv Immunol *109*, 1-44.

1285 Lane, P.J., McConnell, F.M., Withers, D., Gaspal, F., Saini, M., and Anderson, G. (2010).
1286 Lymphoid tissue inducer cells and the evolution of CD4 dependent high-affinity antibody
1287 responses. Prog Mol Biol Transl Sci *92*, 159-174.

1288 Letourneur, F., and Malissen, B. (1989). Derivation of a T cell hybridoma variant deprived of
1289 functional T cell receptor alpha and beta chain transcripts reveals a nonfunctional alpha-mRNA
1290 of BW5147 origin. Eur J Immunol *19*, 2269-2274.

1291 Lockless, S.W., and Ranganathan, R. (1999). Evolutionarily conserved pathways of energetic
1292 connectivity in protein families. Science *286*, 295-299.

1293 Malissen, B., and Bongrand, P. (2015). Early T cell activation: integrating biochemical, structural,
1294 and biophysical cues. Annu Rev Immunol *33*, 539-561.

1295 Parravicini, C., Abbracchio, M.P., Fantucci, P., and Ranghino, G. (2010). Forced unbinding of
1296 GPR17 ligands from wild type and R255I mutant receptor models through a computational
1297 approach. BMC Struct Biol *10*, 8.

1298 Parrish, H.L., Deshpande, N.R., Vasic, J., and Kuhns, M.S. (2016). Functional evidence for TCR-
1299 intrinsic specificity for MHCII. Proc Natl Acad Sci U S A.

1300 Parrish, H.L., Glassman, C.R., Keenen, M.M., Deshpande, N.R., Bronnimann, M.P., and Kuhns,
1301 M.S. (2015). A Transmembrane Domain GGxxG Motif in CD4 Contributes to Its Lck-Independent
1302 Function but Does Not Mediate CD4 Dimerization. PLoS One *10*, e0132333.

1303 Pike, L.J. (2006). Rafts defined: a report on the Keystone Symposium on Lipid Rafts and Cell
1304 Function. J Lipid Res *47*, 1597-1598.

1305 Pond, S.L., Frost, S.D., and Muse, S.V. (2005). HyPhy: hypothesis testing using phylogenies.
1306 Bioinformatics *21*, 676-679.

1307 Pond, S.L.K., and Frost, S.D.W. (2005). Not so different after all: A comparison of methods for
1308 detecting amino acid sites under selection. Mol Biol Evol *22*, 1208-1222.

1309 Popik, W., and Alce, T.M. (2004). CD4 receptor localized to non-raft membrane microdomains
1310 supports HIV-1 entry. Identification of a novel raft localization marker in CD4. J Biol Chem *279*,
1311 704-712.

1312 Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2--approximately maximum-likelihood
1313 trees for large alignments. PLoS One *5*, e9490.

1314 Shah, P., McCandlish, D.M., and Plotkin, J.B. (2015). Contingency and entrenchment in protein
1315 evolution under purifying selection. Proc Natl Acad Sci U S A *112*, E3226-3235.

1316 Sleckman, B.P., Shin, J., Igras, V.E., Collins, T.L., Strominger, J.L., and Burakoff, S.J. (1992).
1317 Disruption of the CD4-p56lck complex is required for rapid internalization of CD4. Proc Natl
1318 Acad Sci U S A *89*, 7566-7570.

1319 Song, Y., Kenworthy, A.K., and Sanders, C.R. (2014). Cholesterol as a co-solvent and a ligand for
1320 membrane proteins. Protein Sci *23*, 1-22.

1321 Stepanek, O., Prabhakar, A.S., Osswald, C., King, C.G., Bulek, A., Naeher, D., Beaufils-Hugot, M.,
1322 Abanto, M.L., Galati, V., Hausmann, B.*, et al.* (2014). Coreceptor scanning by the T cell receptor
1323 provides a mechanism for T cell tolerance. Cell *159*, 333-345.

1324 Teese, M.G., and Langosch, D. (2015). Role of GxxxG Motifs in Transmembrane Domain
1325 Interactions. Biochemistry *54*, 5125-5135.

1326 Turner, J.M., Brodsky, M.H., Irving, B.A., Levin, S.D., Perlmutter, R.M., and Littman, D.R. (1990).
1327 Interaction of the unique N-terminal region of tyrosine kinase p56lck with cytoplasmic domains
1328 of CD4 and CD8 is mediated by cysteine motifs. Cell *60*, 755-765.

1329 Van Roey, K., Gibson, T.J., and Davey, N.E. (2012). Motif switches: decision-making in cell
1330 regulation. Curr Opin Struct Biol *22*, 378-385.

1331 Venkatesh, B., Lee, A.P., Ravi, V., Maurya, A.K., Lian, M.M., Swann, J.B., Ohta, Y., Flajnik, M.F.,
1332 Sutoh, Y., Kasahara, M.*, et al.* (2014). Elephant shark genome provides unique insights into
1333 gnathostome evolution. Nature *505*, 174-179.

1334 Waddell, P.J., and Steel, M.A. (1997). General time-reversible distances with unequal rates
1335 across sites: mixing gamma and inverse Gaussian distributions with invariant sites. Mol
1336 Phylogenet Evol *8*, 398-414.

1337 Weaver, S., Shank, S.D., Spielman, S.J., Li, M., Muse, S.V., and Kosakovsky Pond, S.L. (2018).
1338 Datamonkey 2.0: A Modern Web Application for Characterizing Selective and Other
1339 Evolutionary Processes. Mol Biol Evol *35*, 773-777.

1340 Whitney, D.S., Volkman, B.F., and Prehoda, K.E. (2016). Evolution of a Protein Interaction
1341 Domain Family by Tuning Conformational Flexibility. J Am Chem Soc *138*, 15150-15156.

1342 Willbold, D., and Rosch, P. (1996). Solution Structure of the Human CD4 (403-419) Receptor
1343 Peptide. J Biomed Sci *3*, 435-441.

1344 Wu, H., Kwong, P.D., and Hendrickson, W.A. (1997). Dimeric association and segmental
1345 variability in the structure of human CD4. Nature *387*, 527-530.

1346 Yeung, T., and Grinstein, S. (2007). Lipid signaling and the modulation of surface charge during
1347 phagocytosis. Immunol Rev *219*, 17-36.

Yin, Y., Wang, X.X., and Mariuzza, R.A. (2012). Crystal structure of a complete ternary complex of T-cell receptor, peptide-MHC, and CD4. Proceedings of the National Academy of Sciences of the United States of America *109*, 5405-5410.

Acknowledgements

This work was supported by the National Institutes of Health/National Institute of Allergy and Infections Diseases Grant R01AI101053 (MSK), the Pew Scholars Program in the Biomedical Sciences (MSK), the Cancer Center Support Grant CCSG-CA 023074 for flow cytometry (MSK), and AZ TRIF funds from the BIO5 institute (KVD). We thank Thomas Serwold for critical feedback as well as Dominik Schenten, Joseph Harrison, Piet Maes for critically reading the manuscript.

Author's Contributions

MSK and KVD conceived of the project. MSK directed the research and KVD performed the computational analysis. MSL, PJT, CK, KL and HLP performed the experiments. The manuscript was written by MSK and KVD. All authors contributed to data analysis, discussions, editing, reading and approval of the manuscript.

Competing interest statement: MSK has disclosed an outside interest in Module Therapeutics to the University of Arizona. Conflicts of interest resulting from this interest are being managed by the University of Arizona in accordance with their policies.

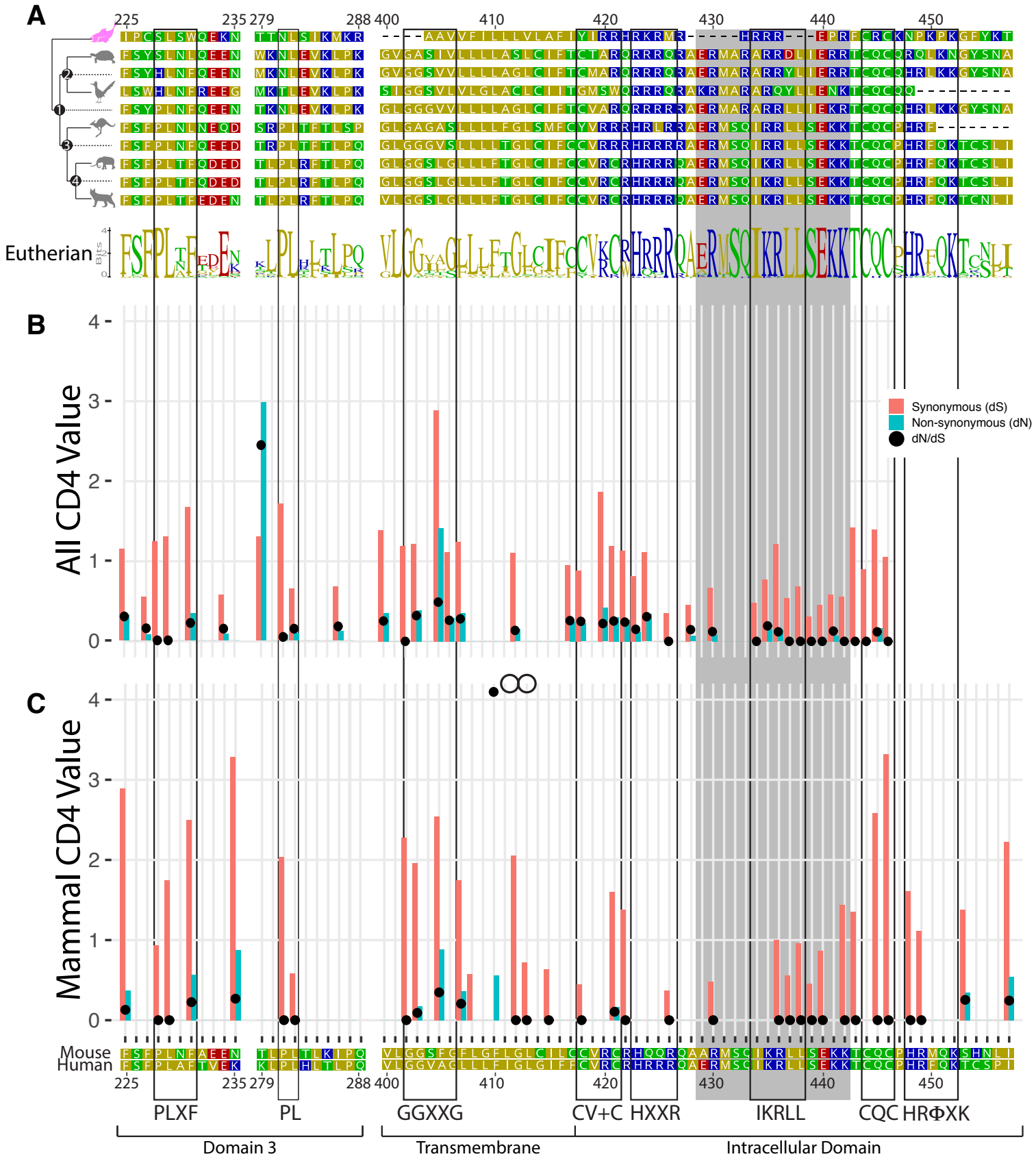

A
225
235 279
288 400
410
420
430
440
450
Eutherian
B
All CD4 Value
Synonymous (dS)
Non-synonymous (dN)
dN/dS
C
Mammal CD4 Value
Mouse
Human
PLXF
PL
GGXXG
CV+C
HXXR
IKRLL
CQC
HRΦXK
Domain 3
Transmembrane
Intracellular Domain

Figure 1

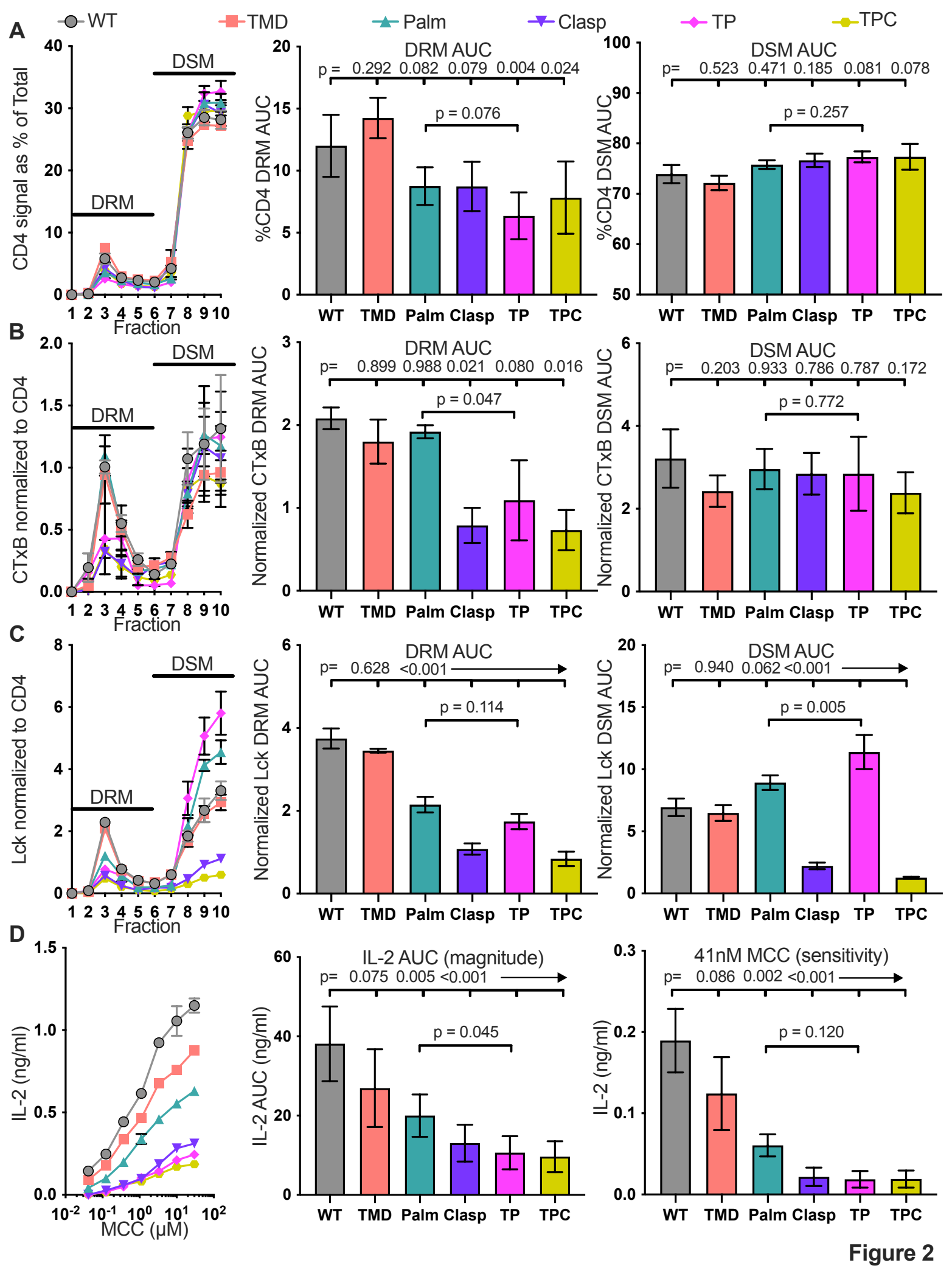

A
WT
TMD
Palm
Clasp
TP
TPC
CD4 signal as % of Total
DRM
DSM
Fraction
DRM AUC
p = 0.292 0.082 0.079 0.004 0.024
p = 0.076
%CD4 DRM AUC
DSM AUC
p= 0.523 0.471 0.185 0.081 0.078
p = 0.257
%CD4 DSM AUC
B
CTxB normalized to CD4
Normalized CTxB DRM AUC
p= 0.899 0.988 0.021 0.080 0.016
p = 0.047
Normalized CTxB DSM AUC
p= 0.203 0.933 0.786 0.787 0.172
p = 0.772
C
Lck normalized to CD4
Normalized Lck DRM AUC
p= 0.628 <0.001
p = 0.114
Normalized Lck DSM AUC
p= 0.940 0.062 <0.001
p = 0.005
D
IL-2 (ng/ml)
MCC (μM)
IL-2 AUC (magnitude)
p= 0.075 0.005 <0.001
p = 0.045
IL-2 AUC (ng/ml)
41nM MCC (sensitivity)
p= 0.086 0.002 <0.001
p = 0.120

Figure 2

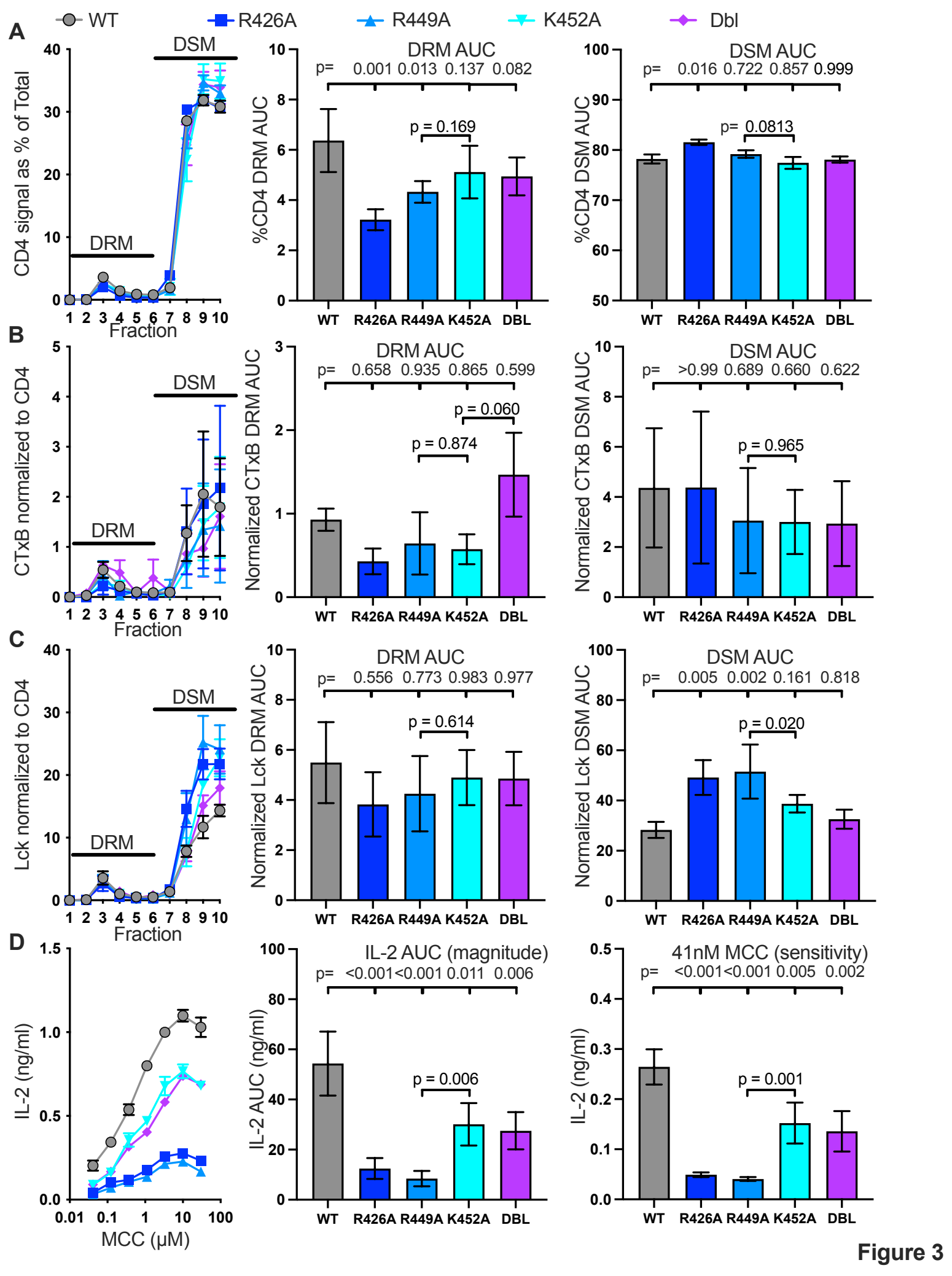

WT
R426A
R449A
K452A
Dbl
A
DSM
DRM
CD4 signal as % of Total
Fraction
DRM AUC
p= 0.001 0.013 0.137 0.082
p = 0.169
%CD4 DRM AUC
DSM AUC
p= 0.016 0.722 0.857 0.999
p= 0.0813
%CD4 DSM AUC
WT R426A R449A K452A DBL
B
CTxB normalized to CD4
DRM AUC
p= 0.658 0.935 0.865 0.599
p = 0.060
p = 0.874
Normalized CTxB DRM AUC
DSM AUC
p= >0.99 0.689 0.660 0.622
p = 0.965
Normalized CTxB DSM AUC
C
Lck normalized to CD4
DRM AUC
p= 0.556 0.773 0.983 0.977
p = 0.614
Normalized Lck DRM AUC
DSM AUC
p= 0.005 0.002 0.161 0.818
p = 0.020
Normalized Lck DSM AUC
D
IL-2 (ng/ml)
MCC (μM)
IL-2 AUC (magnitude)
p= <0.001<0.001 0.011 0.006
p = 0.006
IL-2 AUC (ng/ml)
41nM MCC (sensitivity)
p= <0.001<0.001 0.005 0.002
p = 0.001
IL-2 (ng/ml)

Figure 3

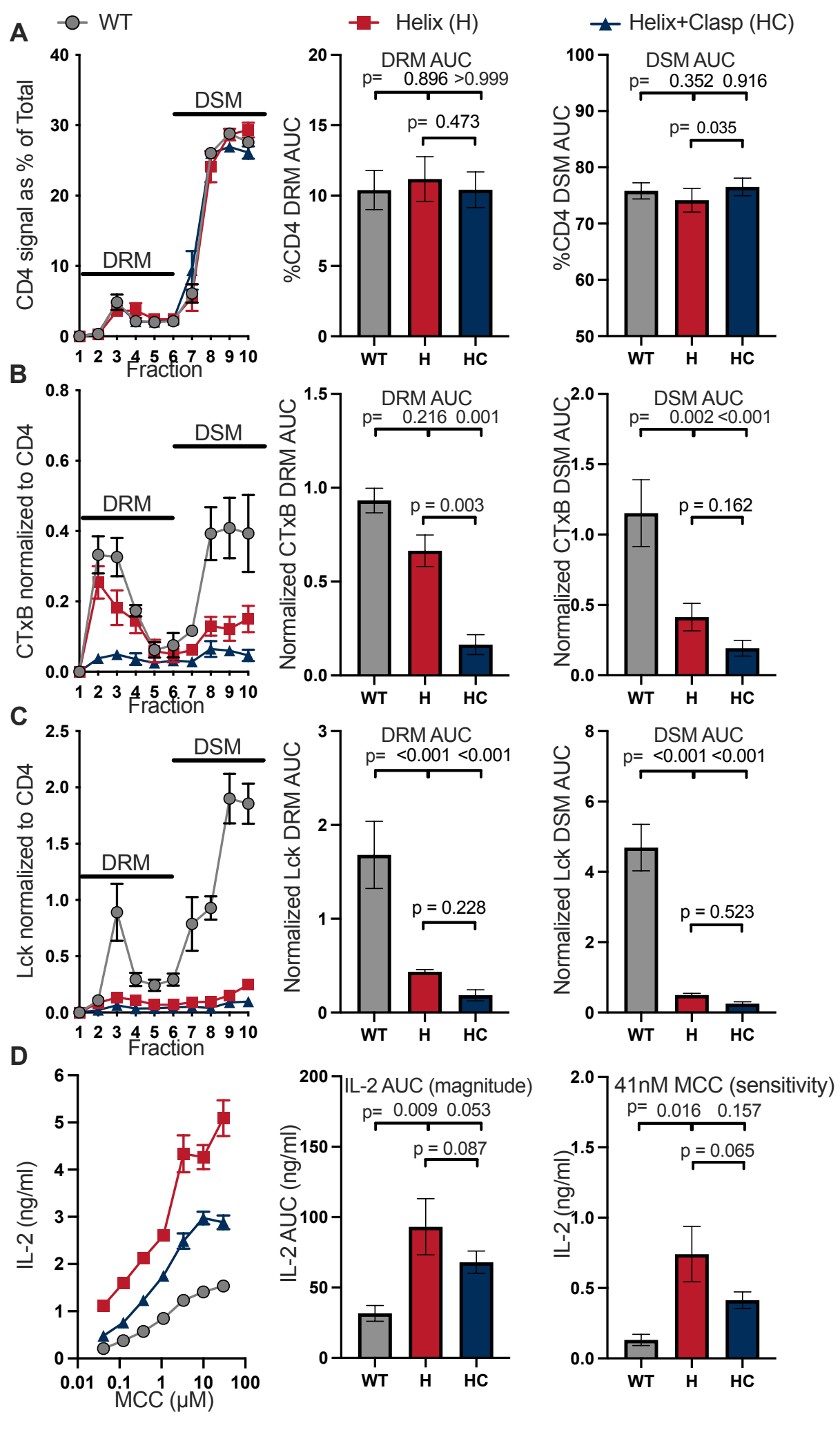

A
WT
Helix (H)
Helix+Clasp (HC)
CD4 signal as % of Total
DRM
DSM
Fraction
DRM AUC
p= 0.896 >0.999
p= 0.473
%CD4 DRM AUC
DSM AUC
p= 0.352 0.916
p= 0.035
%CD4 DSM AUC
WT
H
HC
B
CTxB normalized to CD4
DRM AUC
p= 0.216 0.001
p = 0.003
Normalized CTxB DRM AUC
DSM AUC
p= 0.002 <0.001
p = 0.162
Normalized CTxB DSM AUC
C
Lck normalized to CD4
DRM AUC
p= <0.001 <0.001
p = 0.228
Normalized Lck DRM AUC
DSM AUC
p= <0.001 <0.001
p = 0.523
Normalized Lck DSM AUC
D
IL-2 (ng/ml)
MCC (μM)
IL-2 AUC (magnitude)
p= 0.009 0.053
p = 0.087
IL-2 AUC (ng/ml)
41nM MCC (sensitivity)
p= 0.016 0.157
p = 0.065
IL-2 (ng/ml)

Figure 4

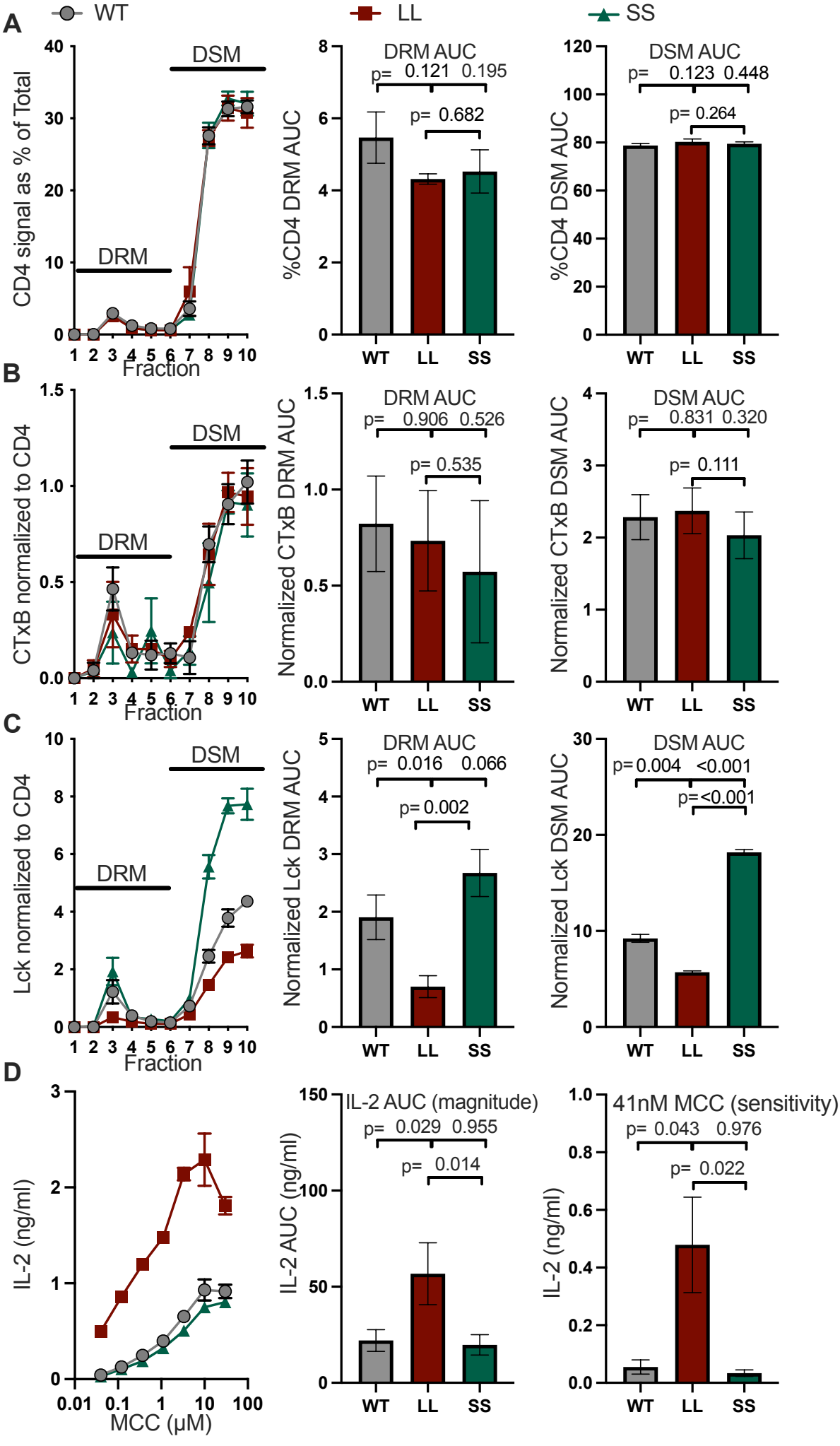

A
WT
LL
SS
CD4 signal as % of Total
DSM
DRM
Fraction
DRM AUC
p= 0.121 0.195
p= 0.682
%CD4 DRM AUC
DSM AUC
p= 0.123 0.448
p= 0.264
%CD4 DSM AUC
B
CTxB normalized to CD4
DRM AUC
p= 0.906 0.526
p= 0.535
Normalized CTxB DRM AUC
DSM AUC
p= 0.831 0.320
p= 0.111
Normalized CTxB DSM AUC
C
Lck normalized to CD4
DRM AUC
p= 0.016 0.066
p= 0.002
Normalized Lck DRM AUC
DSM AUC
p= 0.004 <0.001
p= <0.001
Normalized Lck DSM AUC
D
IL-2 (ng/ml)
MCC (μM)
IL-2 AUC (magnitude)
p= 0.029 0.955
p= 0.014
IL-2 AUC (ng/ml)
41nM MCC (sensitivity)
p= 0.043 0.976
p= 0.022
IL-2 (ng/ml)

Figure 5

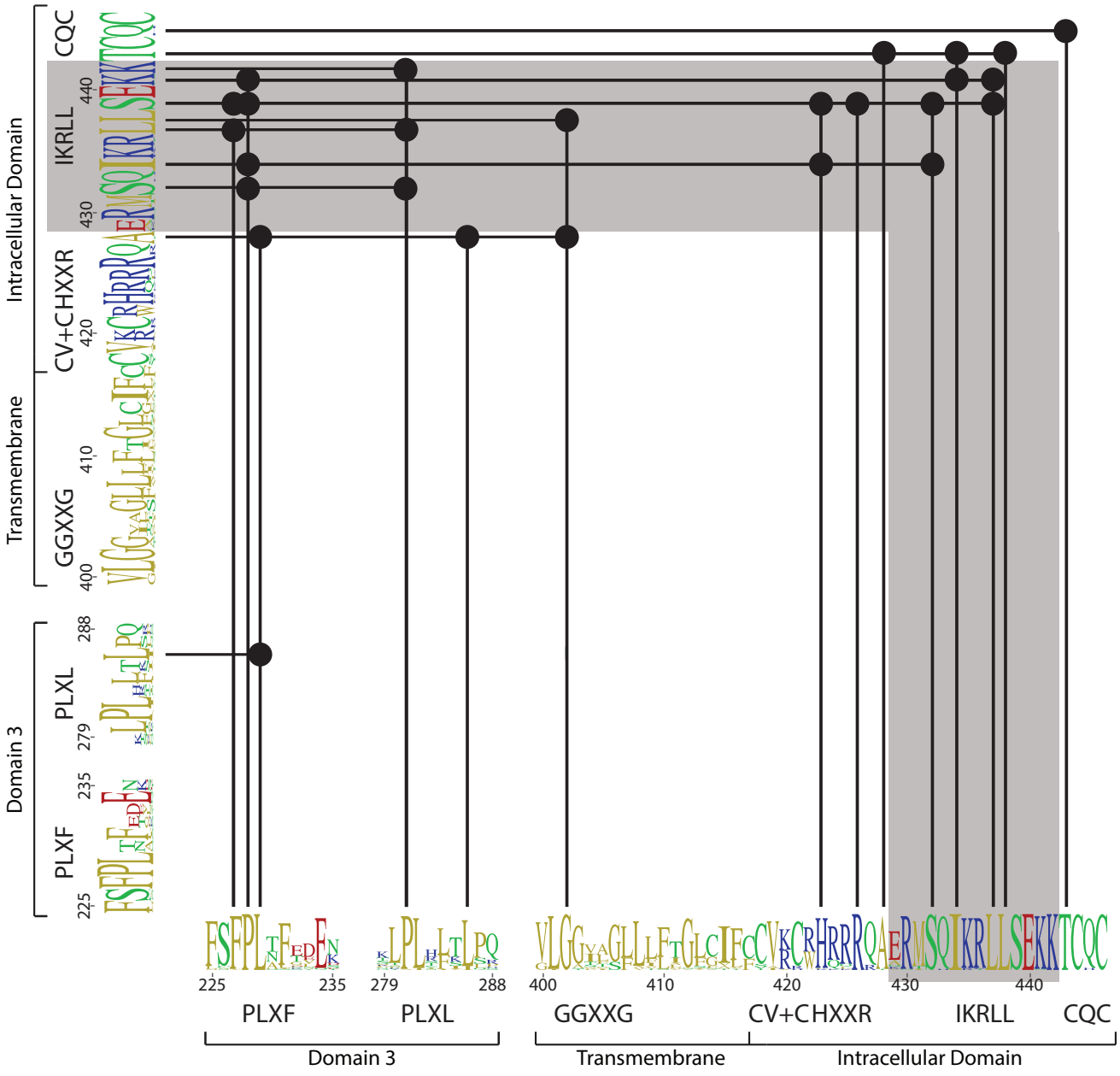

Domain 3
Transmembrane
Intracellular Domain
PLXF
PLXL
GGXXG
CV+CHXXR
IKRLL
CQC
225
235
279
288
400
410
420
430
440

Figure 6

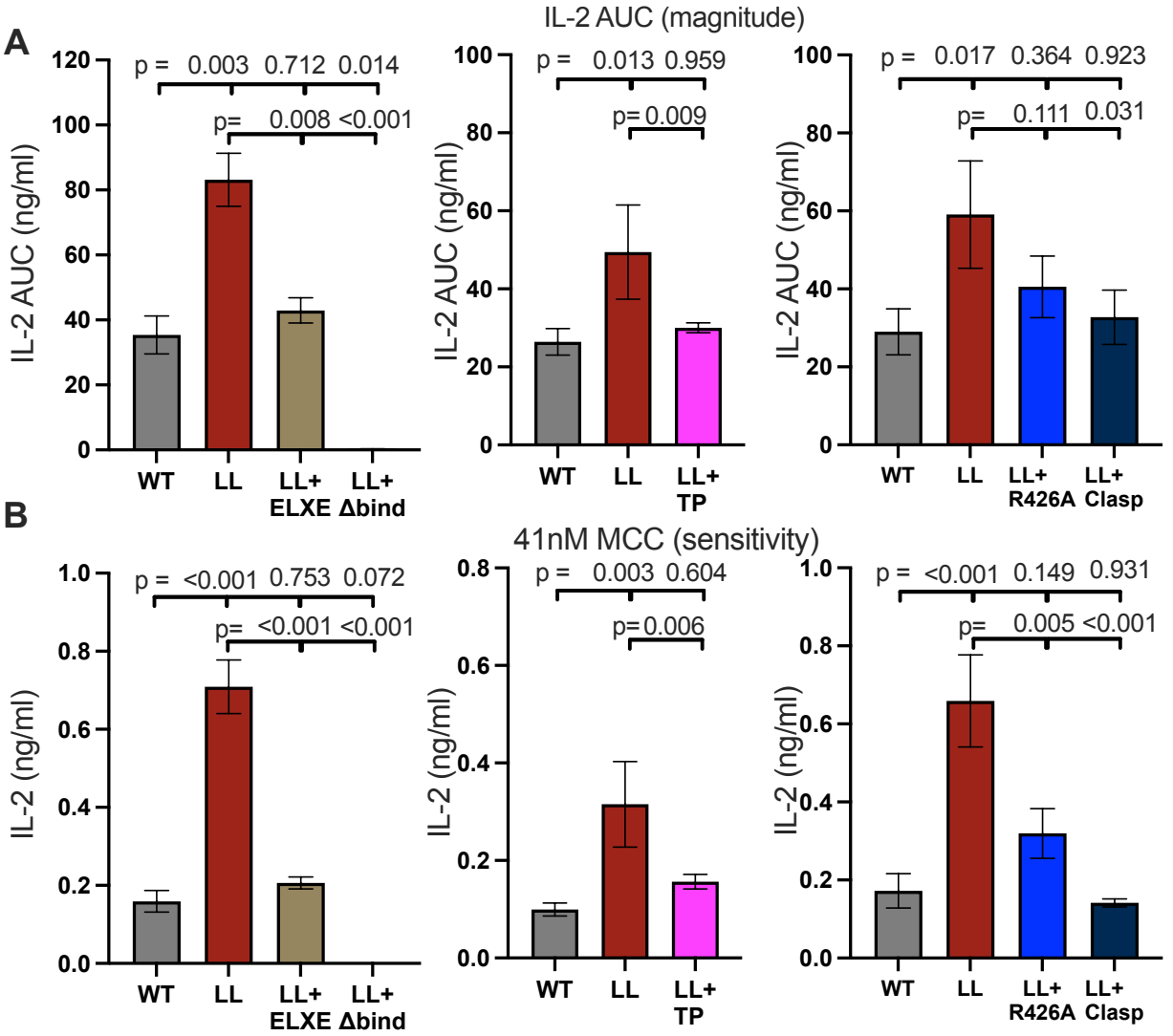

A
IL-2 AUC (magnitude)
IL-2 AUC (ng/ml)
p = 0.003 0.712 0.014
p= 0.008 <0.001
WT
LL
LL+ ELXE
LL+ Δbind
p = 0.013 0.959
p= 0.009
LL+ TP
p = 0.017 0.364 0.923
p= 0.111 0.031
LL+ R426A
LL+ Clasp
B
41nM MCC (sensitivity)
IL-2 (ng/ml)
p = <0.001 0.753 0.072
p= <0.001 <0.001
p = 0.003 0.604
p= 0.006
p = <0.001 0.149 0.931
p= 0.005 <0.001

Figure 7

Figure Legends:

Figure 1. Evolutionary analysis of the CD4 molecule
(A) Reduced representation maximum likelihood phylogenetic tree clusters of CD4 sequences are shown with mouse CD4 numbering (uniprot) used as a reference. Residues are colored based on sidechain polarity with dashes (-) indicating an evolutionary insertion or deletion event. Predicted most recent common ancestor (MRCA) sequences are shown at each node in the tree (Node 1-4). Logo plots of extant eutherian CD4 sequences are aligned at the bottom of the tree in which each stack of letters represents the sequence conservation at that position in the alignment and the height of symbols indicates the relative frequency of each amino acid at that specific position.
(B) Synonymous (dS, red bars) and nonsynonymous (dN, blue bars) substitution rates within the CD4 coding sequence are shown as calculated for all CD4 orthologs included in the initial phylogenetic analysis using the Fixed Effects Likelihood (FEL) method. Only bars for which the likelihood ratio test indicated statistical significance (alpha = 0.1) are shown. Black circles show the ratio of both these values. Codons under diversifying selection have a dN:dS ratio >1 and those under purifying selection have a dN:dS ratio <1.
(C) dS and dN substitution rates are shown as in (B) for the dataset containing only mammalian CD4 sequences to include the region C-terminal of the CQC clasp that is highly variable between phylogenetic clades, and could not be confidently aligned, but is highly conserved in mammals. The codon corresponding to F410 had a dS of 0, leading to a dN/dS ratio of infinity. Human and mouse CD4 sequences are aligned at the bottom for reference.
Boxes are used to highlight motifs discussed in this study, while the grey shading indicates the helix-turn region within the ICD. See also Figure S1.

Figure 2. The GGXXG + CV+C motifs influence CD4 membrane domain localization and function
(A) CD4 signal for each sucrose gradient fraction is shown as a percent of the total CD4 signal detected in all fractions (left). The area under the curve (AUC) is presented for the normalized CD4 signal in the detergent resistant membrane (DRM) fractions (center) and detergent soluble membrane (DSM) fractions (right).
(B) Cholera toxin B (CTxB) signal is shown for each sucrose fraction normalized to the CD4 signal detected in the corresponding fraction (left). The AUC is shown for the normalized CTxB signal in the DRM (center) and DSM fractions (right).
(C) Lck signal is shown for each sucrose fraction normalized to the CD4 signal detected in the corresponding fraction (left). The AUC is shown for the normalized Lck signal in the DRM (center) and DSM fractions (right).
(D) IL-2 production is shown in response to a titration of MCC peptide (left). AUC analysis for the dose response is shown as a measure of the response magnitude (center). The average response to a low dose (41nM) of peptide is shown as a measure of sensitivity (right).
For (A-C) each data point represents the mean +/- SEM for the same three independent experiments (biological replicates). For (D), the dose response represents one of three

experiments showing the mean +/- SEM of triplicate wells (technical replicates). The magnitude and sensitivity data represent the mean +/- SEM of three independent experiments (biological replicates). One-way ANOVA with a Dunnett's posttest for comparisons with WT samples, or a Sidak's posttest for comparisons between selected samples, were performed. See also Figure S2.

Figure 3. The HXXR and HRΦQK motifs influence CD4 membrane domain localization and function
(A) CD4 signal normalized as a percent of the total is shown for each sucrose gradient fraction (left) along with AUC analysis for the DRM (center) and DSM fractions (right).
(B) Cholera toxin B (CTxB) signal normalized to CD4 signal detected is shown for each sucrose fraction (left) along with the AUC analysis for the DRM (center) and DSM fractions (right).
(C) Lck signal normalized to CD4 signal is shown for each sucrose fraction (left) along with the AUC analysis for the DRM (center) and DSM fractions (right).
(D) IL-2 dose response to MCC peptide (left), AUC analysis as a measure of the response magnitude (center), and the average response to a low dose (41nM) of MCC as a measure of sensitivity (right) are shown.
For (A-D) the data are presented as in Figure 2. One-way ANOVA was performed with a Dunnet's posttest for comparisons with WT samples and a Sidak's posttest for comparisons between selected samples. See also Figure S3.

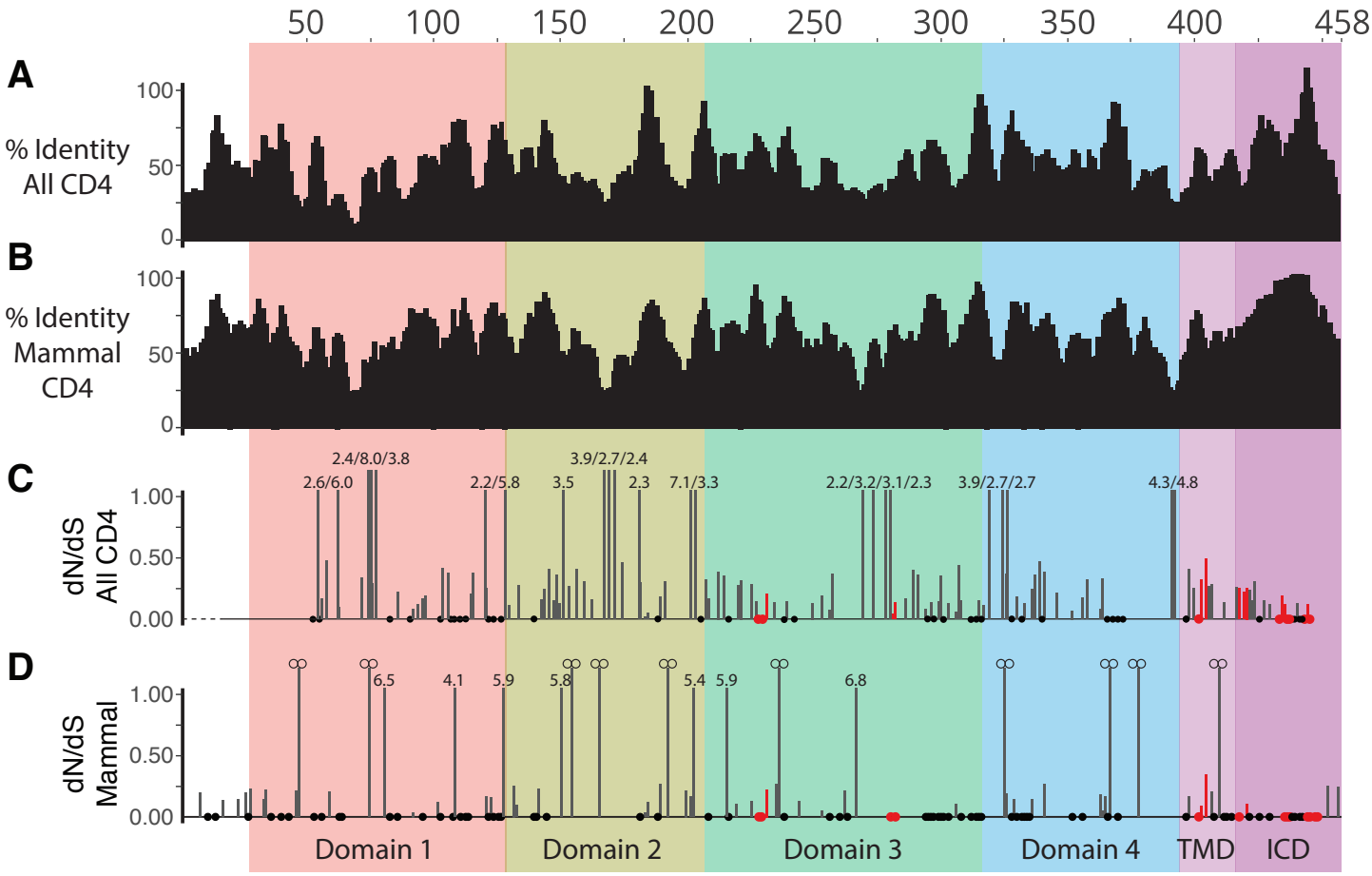
Figure 4. The intracellular helix attenuates response magnitude and sensitivity
(A) CD4 signal normalized as a percent of the total is shown for each sucrose gradient fraction (left) along with AUC analysis for the DRM (center) and DSM fractions (right).
(B) Cholera toxin B (CTxB) signal normalized to CD4 signal detected is shown for each sucrose fraction (left) along with the AUC analysis for the DRM (center) and DSM fractions (right).
(C) Lck signal normalized to CD4 signal is shown for each sucrose fraction (left) along with the AUC analysis for the DRM (center) and DSM fractions (right).
(D) IL-2 dose response to MCC peptide (left), AUC analysis as a measure of the response magnitude (center), and the average response to a low dose (41nM) of MCC as a measure of sensitivity (right) are shown.
For (A-D) the data are presented as in Figure 2. One-way ANOVA was performed with a Dunnett's posttest for comparisons with WT samples and a Sidak's posttest for comparisons between selected samples. See also Figure S4.

Figure 5. The IKRLL motif mediates the inhibitory function of the helix
(A) CD4 signal normalized as a percent of the total is shown for each sucrose gradient fraction (left) along with AUC analysis for the DRM (center) and DSM fractions (right).
(B) Cholera toxin B (CTxB) signal normalized to CD4 signal detected is shown for each sucrose fraction (left) along with the AUC analysis for the DRM (center) and DSM fractions (right).
(C) Lck signal normalized to CD4 signal is shown for each sucrose fraction (left) along with the AUC analysis for the DRM (center) and DSM fractions (right).

(D) IL-2 dose response to MCC peptide (left), AUC analysis as a measure of the response magnitude (center), and the average response to a low dose (41nM) of MCC as a measure of sensitivity (right) are shown.
For (A-D) the data are presented as in Figure 2. One-way ANOVA was performed with a Dunnett's posttest for comparisons with WT samples and a Sidak's posttest for comparisons between selected samples. See also Figure S5.

Figure 6. Covariation matrix
Covarying residues were calculated using MISTIC2. Residues that covary are indicated with a black dot and connected with a solid line. Motifs identified in this study are indicated. The logo plot represents eutherian sequences. The complete MISTIC2 results matrix is available on Dryad (https://doi.org/10.5061/dryad.59zw3r26z).

Figure 7. Coevolving motifs in the ECD and ICD functionally interact
(A) AUC analysis of IL-2 dose responses to MCC peptide are shown as a measure of the response magnitude for the indicated samples.
(B) The average IL-2 response to a low dose (41nM) of MCC is shown as a measure of sensitivity for the indicated samples.
For (A and B) the magnitude and sensitivity data represent the mean +/- SEM of three independent experiments (biological replicates) for which triplicate measurements were performed (technical replicates). One-way ANOVA was performed with a Dunnett's posttest for comparisons with WT samples, and a Sidak's posttest for comparisons between selected samples. Individual graphs indicate experiments that were performed with cell lines generated at the same time. See also Figure S6.


50
100
150
200
250
300
350
400
458
A
% Identity All CD4
100
50
0
B
% Identity Mammal CD4
100
50
0
C
dN/dS All CD4
1.00
0.50
0.00
2.4/8.0/3.8
2.6/6.0
2.2/5.8
3.9/2.7/2.4
3.5
2.3
7.1/3.3
2.2/3.2/3.1/2.3
3.9/2.7/2.7
4.3/4.8
D
dN/dS Mammal
1.00
0.50
0.00
6.5
4.1
5.9
5.8
5.4
5.9
6.8
Domain 1
Domain 2
Domain 3
Domain 4
TMD
ICD

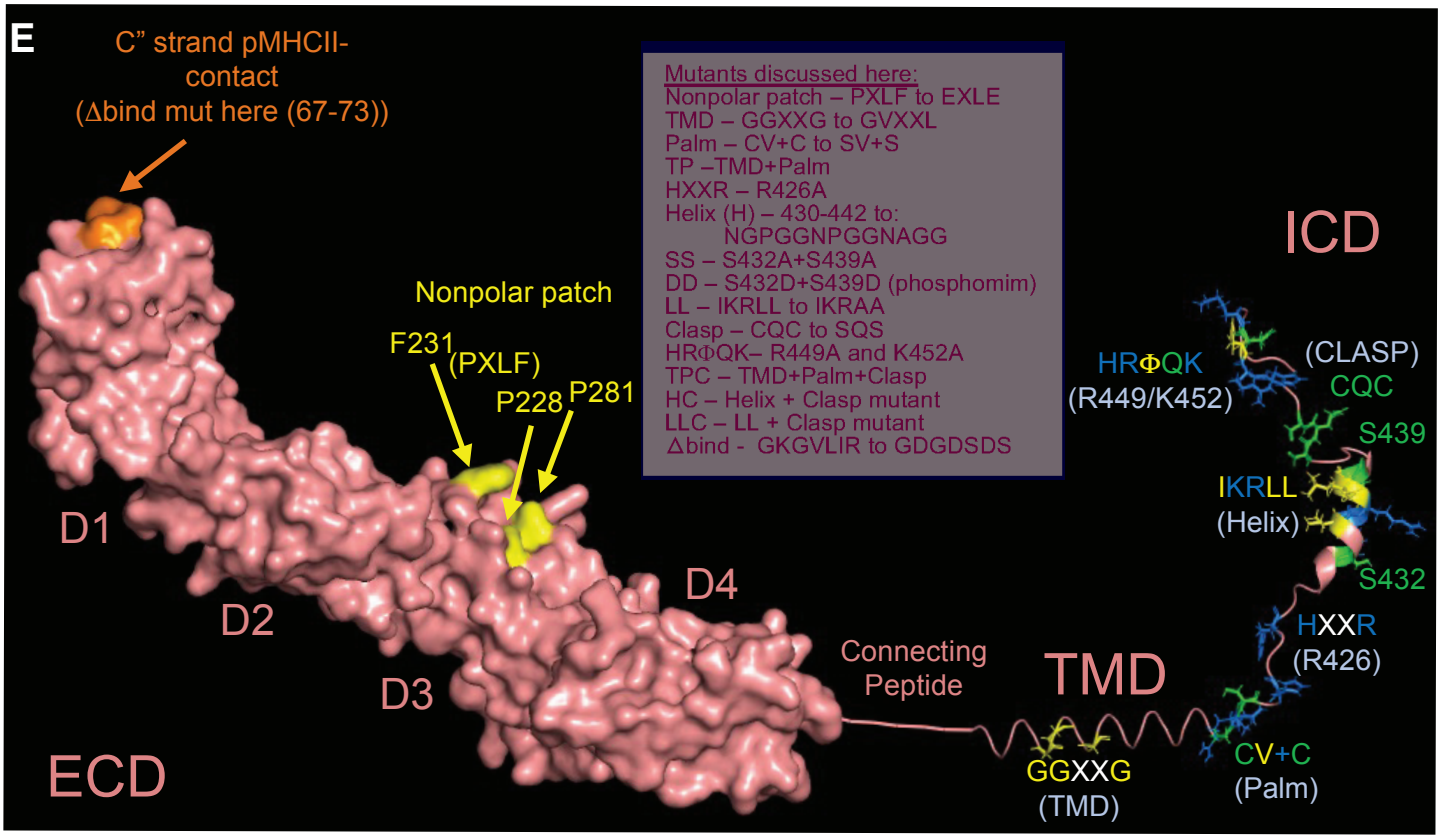

E
C" strand pMHCII-contact
(Δbind mut here (67-73))
Mutants discussed here:
Nonpolar patch – PXLF to EXLE
TMD – GGXXG to GVXXL
Palm – CV+C to SV+S
TP –TMD+Palm
HXXR – R426A
Helix (H) – 430-442 to:
NGPGGNPGGNAGG
SS – S432A+S439A
DD – S432D+S439D (phosphomim)
LL – IKRLL to IKRAA
Clasp – CQC to SQS
HRΦQK– R449A and K452A
TPC – TMD+Palm+Clasp
HC – Helix + Clasp mutant
LLC – LL + Clasp mutant
Δbind - GKGVLIR to GDGDSDS
Nonpolar patch
F231
(PXLF)
P228
P281
D1
D2
D3
D4
ECD
Connecting Peptide
TMD
GGXXG
(TMD)
CV+C
(Palm)
HXXR
(R426)
S432
IKRLL
(Helix)
S439
CQC
(CLASP)
HRΦQK
(R449/K452)
ICD

Figure S1

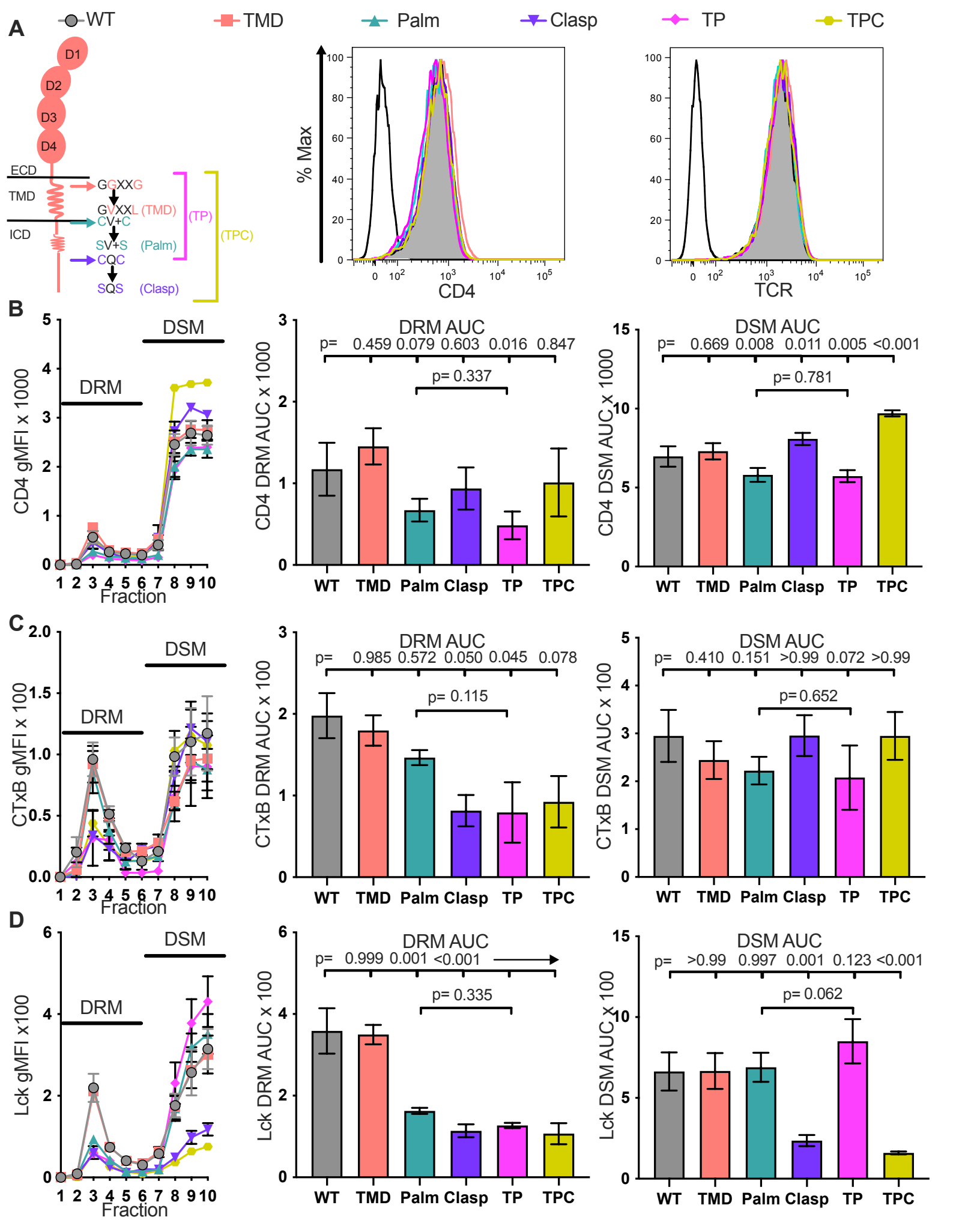

A
WT
TMD
Palm
Clasp
TP
TPC
D1
D2
D3
D4
ECD
TMD
ICD
GGXXG
GVXXL (TMD)
CV+C
SV+S (Palm)
CQC
SQS (Clasp)
(TP)
(TPC)
% Max
CD4
TCR
B
DSM
DRM
CD4 gMFI x 1000
Fraction
DRM AUC
p= 0.459 0.079 0.603 0.016 0.847
p= 0.337
CD4 DRM AUC x 1000
DSM AUC
p= 0.669 0.008 0.011 0.005 <0.001
p= 0.781
CD4 DSM AUC x 1000
WT TMD Palm Clasp TP TPC
C
CTxB gMFI x 100
DRM AUC
p= 0.985 0.572 0.050 0.045 0.078
p= 0.115
CTxB DRM AUC x 100
DSM AUC
p= 0.410 0.151 >0.99 0.072 >0.99
p= 0.652
CTxB DSM AUC x 100
D
Lck gMFI x100
DRM AUC
p= 0.999 0.001 <0.001
p= 0.335
Lck DRM AUC x 100
DSM AUC
p= >0.99 0.997 0.001 0.123 <0.001
p= 0.062
Lck DSM AUC x 100

Figure S2

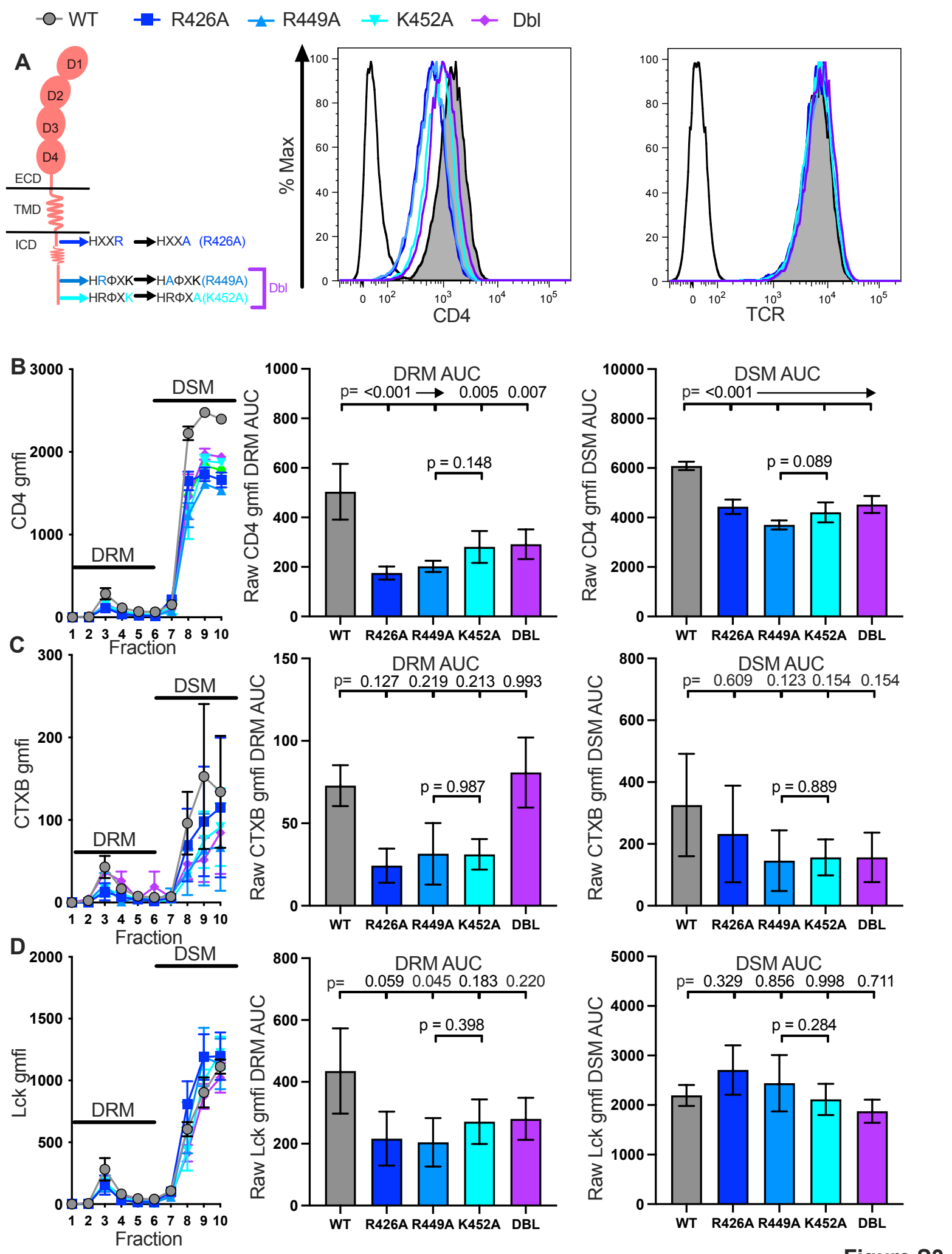

WT
R426A
R449A
K452A
Dbl
A
D1
D2
D3
D4
ECD
TMD
ICD
HXXR
HXXA (R426A)
HRΦXK
HAΦXK(R449A)
HRΦXK
HRΦXA(K452A)
Dbl
% Max
CD4
TCR
B
DSM
DRM
CD4 gmfi
Fraction
Raw CD4 gmfi DRM AUC
DRM AUC
p= <0.001 0.005 0.007
p = 0.148
Raw CD4 gmfi DSM AUC
DSM AUC
p= <0.001
p = 0.089
WT R426A R449A K452A DBL
C
CTXB gmfi
Raw CTXB gmfi DRM AUC
DRM AUC
p= 0.127 0.219 0.213 0.993
p = 0.987
Raw CTXB gmfi DSM AUC
DSM AUC
p= 0.609 0.123 0.154 0.154
p = 0.889
D
Lck gmfi
Raw Lck gmfi DRM AUC
DRM AUC
p= 0.059 0.045 0.183 0.220
p = 0.398
Raw Lck gmfi DSM AUC
DSM AUC
p= 0.329 0.856 0.998 0.711
p = 0.284

Figure S3

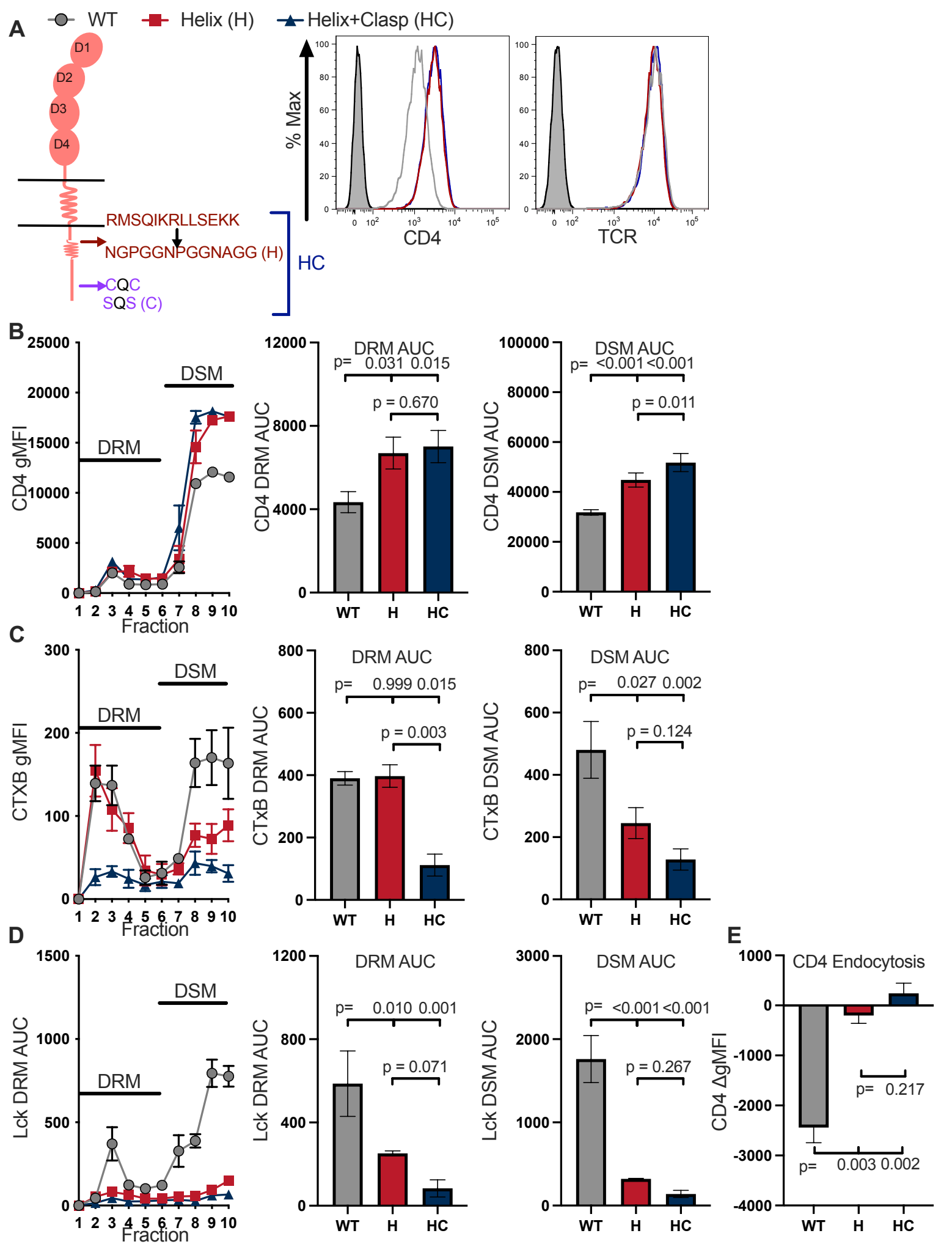

A
WT
Helix (H)
Helix+Clasp (HC)
D1
D2
D3
D4
RMSQIKRLLSEKK
NGPGGNPGGNAGG (H)
CQC
SQS (C)
HC
% Max
CD4
TCR
B
DSM
DRM
CD4 gMFI
Fraction
DRM AUC
p= 0.031 0.015
p = 0.670
CD4 DRM AUC
DSM AUC
p= <0.001 <0.001
p = 0.011
CD4 DSM AUC
WT
H
HC
C
CTXB gMFI
p= 0.999 0.015
p = 0.003
CTxB DRM AUC
p= 0.027 0.002
p = 0.124
CTxB DSM AUC
D
Lck DRM AUC
p= 0.010 0.001
p = 0.071
Lck DSM AUC
p= <0.001 <0.001
p = 0.267
E
CD4 Endocytosis
CD4 ΔgMFI
p= 0.217
p= 0.003 0.002

Figure S4

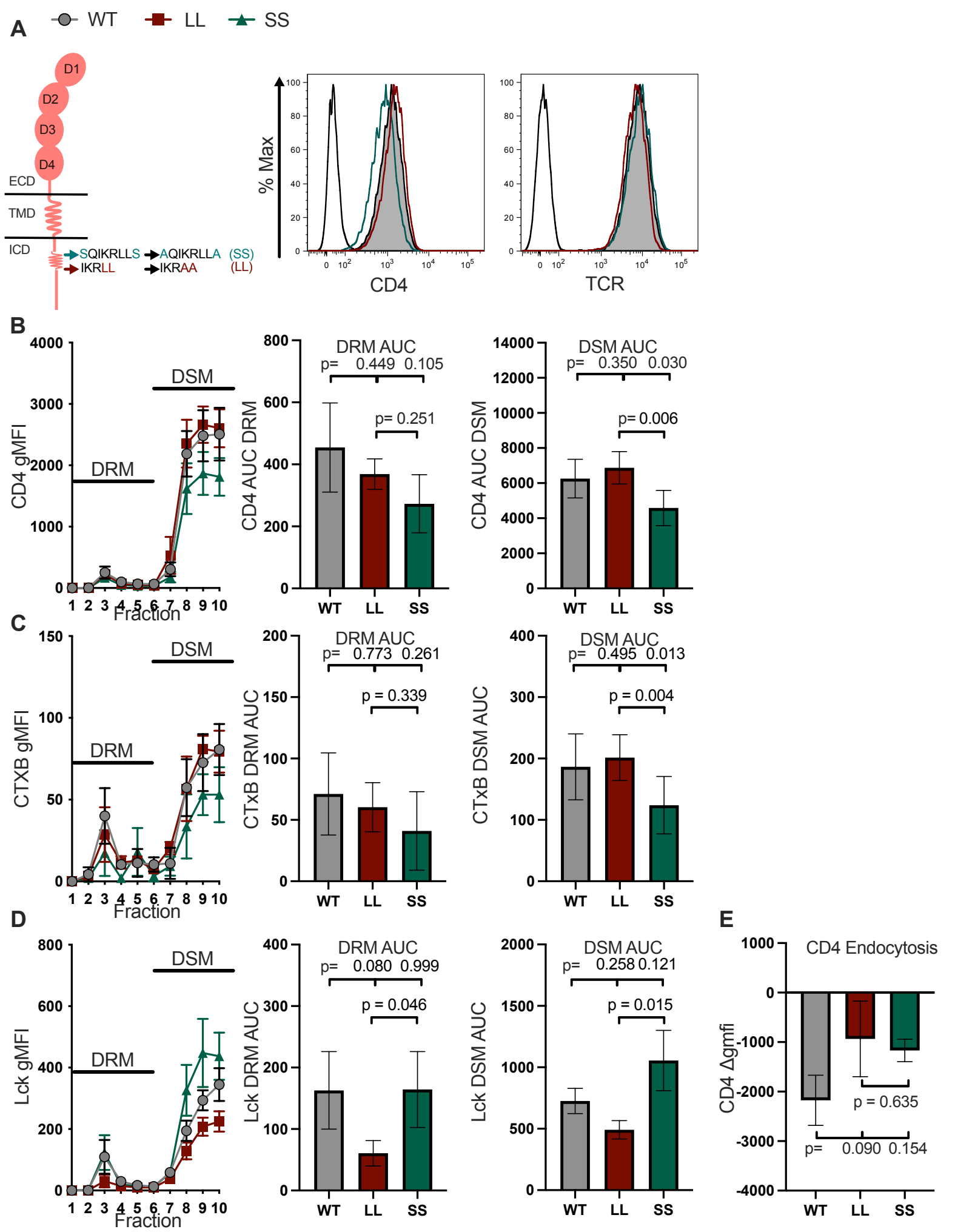

A
WT
LL
SS
D1
D2
D3
D4
ECD
TMD
ICD
SQIKRLLS
AQIKRLLA (SS)
IKRLL
IKRAA (LL)
% Max
CD4
TCR
B
CD4 gMFI
DRM
DSM
Fraction
CD4 AUC DRM
DRM AUC
p= 0.449 0.105
p= 0.251
CD4 AUC DSM
DSM AUC
p= 0.350 0.030
p= 0.006
WT
LL
SS
C
CTXB gMFI
CTxB DRM AUC
DRM AUC
p= 0.773 0.261
p = 0.339
CTxB DSM AUC
DSM AUC
p= 0.495 0.013
p = 0.004
D
Lck gMFI
Lck DRM AUC
DRM AUC
p= 0.080 0.999
p = 0.046
Lck DSM AUC
DSM AUC
p= 0.258 0.121
p = 0.015
E
CD4 Endocytosis
CD4 Δgmfi
p = 0.635
p= 0.090 0.154

Figure S5

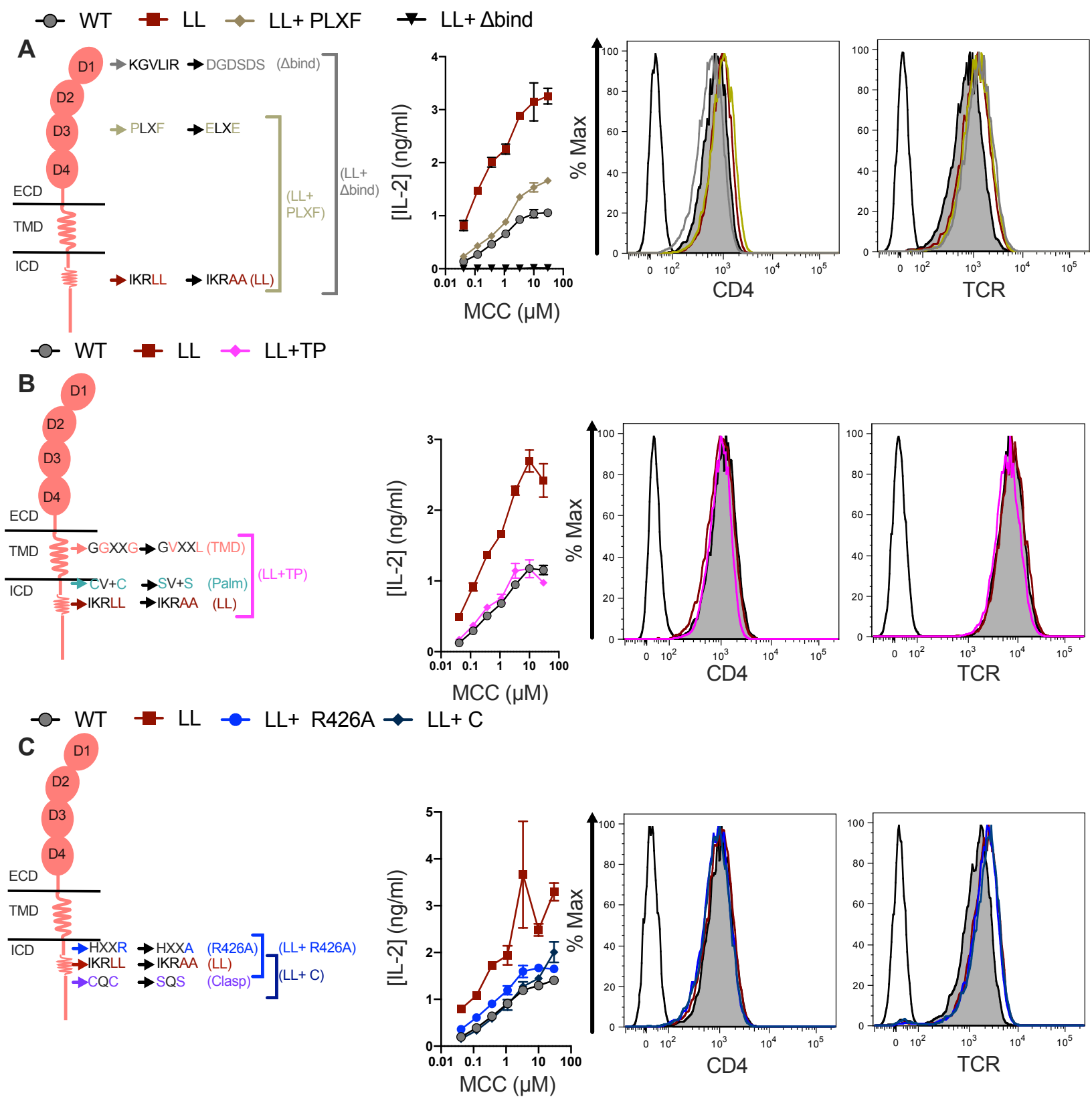

WT
LL
LL+ PLXF
LL+ Δbind
A
D1
D2
D3
D4
ECD
TMD
ICD
KGVLIR
DGDSDS
(Δbind)
PLXF
ELXE
IKRLL
IKRAA (LL)
(LL+ PLXF)
(LL+ Δbind)
[IL-2] (ng/ml)
MCC (μM)
% Max
CD4
TCR
WT
LL
LL+TP
B
GGXXG
GVXXL (TMD)
CV+C
SV+S
(Palm)
IKRLL
IKRAA
(LL)
(LL+TP)
WT
LL
LL+ R426A
LL+ C
C
HXXR
HXXA
(R426A)
IKRLL
IKRAA
(LL)
CQC
SQS
(Clasp)
(LL+ R426A)
(LL+ C)

Figure S6

Figure Legends:

Figure S1. Conservation and purifying selection of the CD4 molecule
(A) Mean pairwise sequence identity over all pairs in the alignment column was calculated using a sliding window of size 5. The alignment was based on 99 CD4 orthologs and still contains the sequences downstream of the CQC clasp. These residues were later removed based on low sequence conservation.
(B) As in (A) but including only mammalian CD4 orthologs.
(C) The Fixed Effects Likelihood (FEL) model was used to estimate non-synonymous (dN) and synonymous (dS) substitution rates on a per-site basis as an indication of evolutionary selection. Bars show the dN/dS ratio of both these values. Only bars for which the likelihood ratio test indicated statistical significance ($P < 0.1$) are shown. dN/dS ratios below 1 are indicative of purifying selection, while a ratio great than 1 are under positive selection. For visualization purposes, we focus on values lower than 1. For codons under positive selection (i.e, > 1), the value indicates the dN/dS ratio. An infinity symbol was used to indicate those codons for which dS = 0. Circles have a dN/dS ration of 0. Red highlighted codons are located within motifs identified within this study. CD4 domains are indicated using color. Numbering according to mouse CD4.
(E) A theoretical structural model to show the relative location of the motifs discussed here. The surface rendered ECD (pdb 1WIQ), joined with a connecting peptide and TMD (built using the PyMol Molecular Graphics system), and ICD (pdb 1Q68).
See also Figure 1.

Figure S2. Analysis of TMD, Palm, Clasp, TP, and TPC mutant cell lines
(A) Cartoon of CD4 highlighting the location and nature of the mutations tested (left). Flow cytometry analysis of CD4 (center) and TCR expression (right) on $58\alpha^-\beta^-$ cells.
(B) Raw CD4 signal data is shown for each sucrose fraction (left). AUC is shown for the DRM fractions (center) and DSM fractions (right).
(C) Raw cholera toxin B (CTxB) signal data is shown for each sucrose fraction (left). AUC is shown for the DRM (center) and DSM fractions (right).
(D) Raw Lck signal data is shown for each sucrose fraction (left). AUC is shown for the DRM (center) and DSM fractions (right).
For (B-D) each data point represents the mean +/- SEM for the same three independent experiments (biological replicates). One-way ANOVA was performed with a Dunnett's posttest for comparisons with WT samples, and a Sidak's posttest for comparisons between selected samples. See also Figure 2.

Figure S3. Analysis of R426A, R449A, K452A, and Dbl mutant cell lines
(A) Cartoon of CD4 highlighting the location and nature of the mutations tested (left). Flow cytometry analysis of CD4 (center) and TCR expression (right) on $58\alpha^-\beta^-$ cells.
(B) Raw CD4 signal data is shown for each sucrose fraction (left). AUC is shown for the DRM fractions (center) and DSM fractions (right).
(C) Raw cholera toxin B (CTxB) signal data is shown for each sucrose fraction (left). AUC is shown for the DRM (center) and DSM fractions (right).
(D) Raw Lck signal data is shown for each sucrose fraction (left). AUC is shown for the DRM (center) and DSM fractions (right).

For (B-D) each data point represents the mean +/- SEM for the same three independent experiments (biological replicates). One-way ANOVA was performed with a Dunnett's posttest for comparisons with WT samples, and a Sidak's posttest for comparisons between selected samples. See also Figure 3.

Figure S4. Analysis of H and HC mutant cell lines
(A) Cartoon of CD4 highlighting the location and nature of the mutations tested (left). Flow cytometry analysis of CD4 (center) and TCR expression (right) on $58\alpha^-\beta^-$ cells.
(B) Raw CD4 signal data is shown for each sucrose fraction (left). AUC is shown for the DRM fractions (center) and DSM fractions (right).
(C) Raw cholera toxin B (CTxB) signal data is shown for each sucrose fraction (left). AUC is shown for the DRM (center) and DSM fractions (right).
(D) Raw Lck signal data is shown for each sucrose fraction (left). AUC is shown for the DRM (center) and DSM fractions (right).
(E) CD4 endocytosis after pMHCII engagement is shown for the indicated cell lines after 16 hours coculture with APCs in the presence of 10μM MCC peptide. The change in CD4 gMFI, as measured by flow cytometry, is shown for each cell line relative to an equivalent sample cultured with APCs in the absence of MCC peptide.
For (B-E) each data point represents the mean +/- SEM for the same three independent experiments (biological replicates). For E, endocytosis measurements were performed in triplicate (technical replicates) for each experiment. One-way ANOVA was performed with a Dunnett's posttest for comparisons with WT samples, and a Sidak's posttest for comparisons between selected samples. See also Figure 4.

Figure S5. Analysis of LL and SS cell lines
(A) Cartoon of CD4 highlighting the location and nature of the mutations tested (left). Flow cytometry analysis of CD4 (center) and TCR expression (right) on $58\alpha^-\beta^-$ cells.
(B) Raw CD4 signal is shown for each sucrose fraction (left). AUC is shown for the DRM fractions (center) and DSM fractions (right).
(C) Raw cholera toxin B (CTxB) signal is shown for each sucrose fraction (left). AUC is shown for the DRM (center) and DSM fractions (right).
(D) Raw Lck signal is shown for each sucrose fraction (left). AUC is shown for the DRM (center) and DSM fractions (right).
(E) CD4 endocytosis after pMHCII engagement is shown for the indicated cell lines after 16 hours coculture with APCs in the presence of 10μM MCC peptide. The change in CD4 gMFI, as measured by flow cytometry, is shown for each cell line relative to an equivalent sample cultured with APCs in the absence of MCC peptide.
For (B-E) each data point represents the mean +/- SEM for the same three independent experiments (biological replicates). For E, endocytosis measurements were performed in triplicate (technical replicates) for each experiment. One-way ANOVA was performed with a Dunnett's posttest for comparisons with WT samples, and a Sidak's posttest for comparisons between selected samples. See also Figure 5.

Figure S6. Analysis of CD4 mutants of coevolving motifs
(A-D) Cartoon of CD4 highlighting the location and nature of the mutations tested (left). Representative IL-2 dose responses to MCC peptide for one of three independent

experiments with each dataset representing the mean +/- SEM of triplicates (center). Flow cytometry analysis of CD4 and TCR expression on 58$\alpha^-\beta^-$ cells (right). See also Figure 7.