T cell receptor cross-reactivity between gliadin and bacterial peptides in celiac disease

Jan Petersen[1,2], Laura Ciacchi[1,2], Mai T. Tran[1], Khai-Lee Loh[1], Yvonne Kooy-Winkelaar[3], Nathan P. Croft[1], Melinda Y. Hardy[4,5], Zhenjun Chen[6], James McCluskey[6], Robert P. Anderson[7], Anthony W. Purcell[1], Jason A. Tye-Din[4,5,8], Frits Koning[3], Hugh H. Reid[1,2*] & Jamie Rossjohn[1,2,9]*

[1]Infection and Immunity Program and The Department of Biochemistry and Molecular Biology, Biomedicine Discovery Institute Monash University, Clayton, Victoria 3800, Australia.

[2]Australian Research Council Centre of Excellence in Advanced Molecular Imaging, Monash University, Clayton, Victoria 3800, Australia.

[3]Department of Immunohematology and Blood Transfusion, Leiden University Medical Center, Leiden 2333 ZA, The Netherlands.

[4]The Walter and Eliza Hall Institute of Medical Research, 1G Royal Parade, Parkville, Victoria 3052, Australia.

[5]Department of Medical Biology, The University of Melbourne, Parkville, Victoria 3052, Australia.

[6]Department of Microbiology & Immunology, Peter Doherty Institute for Infection and Immunity, University of Melbourne, Melbourne, Victoria 3010, Australia

[7]ImmusanT, Inc., One Kendall Square, Suite B2004, Cambridge, MA 02139, USA

[8]Department of Gastroenterology, The Royal Melbourne Hospital, Parkville, Victoria 3052, Australia.

[9]Institute of Infection and Immunity, Cardiff University School of Medicine, Heath Park, Cardiff CF14 4XN, UK.

[*] joint senior & corresponding authors. hugh.reid@monash.edu and jamie.rossjohn@monash.edu

Running title: Molecular mimicry in Celiac Disease

The *Human Leukocyte Antigen* (HLA) *locus* is strongly associated with T cell mediated autoimmune disorders. HLA-DQ2.5-mediated celiac disease (CeD) is triggered by ingestion of gluten, although the relative role of genetic and environmental risk factors in CeD is unclear. Here we identify microbially-derived mimics of gliadin epitopes and a parental bacterial protein that is naturally processed by antigen presenting cells and activated gliadin-reactive HLA-DQ2.5 restricted T cells derived from CeD patients. Crystal structures of T cell receptors (TCRs) in complex with HLA-DQ2.5 bound to two distinct bacterial peptides demonstrate that molecular mimicry underpins cross-reactivity towards the gliadin epitopes. Accordingly, gliadin reactive T cells involved in CeD pathogenesis cross-react with ubiquitous bacterial peptides, thereby suggesting microbial exposure as a potential environmental factor in CeD.

Introduction

The ability of T cells to distinguish between self- and non-self is determined by T cell receptors (TCRs) on the surface of T cells recognising peptides bound to Human Leukocyte Antigen (HLA) molecules. However, failure of self/non-self discrimination may cause aberrant T cell reactivity against self-peptides and the manifestation of immune-mediated inflammatory diseases (IMID). However, the factors leading to T cell autoimmunity remain obscure.

The *Human Leukocyte Antigen Class II* (*HLA-II*) *locus* represents an important genetic determinant in IMIDs[1]. Nevertheless, the molecular mechanism underpinning this genetic linkage for the vast majority of IMIDs are unclear. The HLA-bound peptides that precipitate autoimmunity are often unknown and it is unclear why tolerance is broken[2]. Moreover, there is a poor understanding of the T cell repertoire directed towards autoreactive peptide-HLA complexes. Notably, the non-HLA linked genetic risk is generally thought to be multi-factorial and shared between different IMIDs, with several lines of evidence implying that environmental factors are likely to play an important role as well. Disease associations with major environmental risk factors are much more diffuse and include the exposure to toxic chemicals [3], the dysregulation of the gut microbiota [4] and certain infections [5]. However, the relative contributions and mechanisms governing environmental and genetic risk factors in autoimmunity and autoimmune-like disorders remain unclear.

To begin to address these central questions regarding IMIDs, we have investigated the molecular mechanisms of the $CD4^+$ T cell response in Celiac Disease (CeD), as the nature of the antigens that triggers the disease is well established [6-8]. CeD is a T cell mediated IMID that is caused by an environmental antigen, dietary gluten. CeD is essentially restricted to genetically predisposed

individuals, namely those who are HLA-DQ2.5^{+} and/or HLA-DQ8^{+}. Approximately 95% of CeD patients carry HLA-DQ2.5 (alleles *DQA1*05:01* and *DQB1* 02:01*), and the majority of HLA-DQ2.5^{-} patients express HLA-DQ8 (alleles *DQA1*03:01* and *DQB1*03:02*) [6]. In CeD, gluten peptide deamidation by tissue transglutaminase 2 (TG2) strengthens binding of epitopes to these disease-associated HLA molecules [9,10], thereby increasing both antigen presentation, and the recognition by gluten-specific CD4^{+} T cells [11-14]. T cell activity in HLA-DQ2.5-associated CeD is targeted towards gluten peptides originating from wheat, rye or barley. In relation to wheat, the immunodominant deamidated gluten peptide encompasses two overlapping T cell determinants in wheat gliadin (DQ2.5-glia-α1a, PFPQPELPY, and DQ2.5-glia-α2: PQPELPYPQ). HLA-DQ2.5-mediated CeD is characterised by biased TCR usage [12-15]. T cells expressing TRAV26-1 and TRBV7-2 predominate in the HLA-DQ2.5-glia-α2 specific T cell response while HLA-DQ2.5-glia-α1a restricted TCRs show biased usage of TRBV29-1 genes. The HLA-DQ2.5-glia-α2 specific T cell response is also characterised by an arginine residue in the CDR3 loop, which plays a key role in TCR recognition. Crystal structures of a TRAV4/TRBV20-1^{+} TCR-DQ2.5-glia-α1a complex and TRAV26-1/TRBV7-2^{+} TCR-HLA-DQ2.5-glia-α2 complexes provided a molecular mechanism underpinning these epitope specificities and the HLA-DQ2.5-glia-α2 restricted TCR bias [14].

However, while the role of HLA-DQ2 and HLA-DQ8 in the presentation of gluten epitopes is clear, the presence of these *HLA* alleles is insufficient to cause disease, as the *HLA* penetrance for this disease is no more than 3% [6], and age of onset is at a median of 3 years and rarely before the second year of life [16]. Thus, other genetic and environmental factors must contribute to disease development. While epidemiological studies have linked environmental events such as infections with the onset of CeD these cannot establish causality so there is a strong need to understand the mechanistic basis for these observations [17-19]. An emerging view is that environmental factors can modify the gut microbiota composition and/or function in genetically susceptible hosts contributing to dysfunctional microbe-host interactions that promote CeD [21]. Reported protein sequence homology between gliadin and adenovirus raised the possibility of molecular mimicry as a contributor to CeD pathogenesis in 1984 but without data supporting an effect in disease relevant T cells this theory was largely disregarded [20,21]. Here we identify and characterise a number of mimics of HLA-DQ2.5-restricted gliadin determinants derived from common environmental bacteria, and show that they activate disease-relevant, gliadin-reactive T cells isolated from CeD patients. Using structural biology, we show that molecular mimicry underpins this cross-reactive TCR response. Accordingly, our results suggest TCR cross-reactivity between gliadin and microbial peptides in HLA-DQ2.5^{+} individuals is a plausible biological mechanism that supports a potential pathogenic link between some environmentally-derived bacteria and CeD.

Results

Identification of microbial mimic peptides of immunodominant CeD epitopes

To establish whether gliadin reactive T cells could cross-react with structurally similar antigens from bacteria, we searched for microbial peptide sequences that could act as molecular mimics of immunodominant CeD-associated deamidated gliadin epitopes. To reduce the extent of false positive hits from a sequence-only based search, we undertook a structurally-guided database search, using constraints imposed by knowledge of the TCR-HLA-DQ2.5-gliadin crystal structures. Namely, (i) the HLA DQ-2.5 anchor residues at positions P4 (mimics of HLA-DQ2.5 glia-α1a and the related DQ2.5-glia-ω1) and P6 (HLA-DQ2.5 glia-α2 mimics) were fixed to a negatively charged residue (Glu or Asp); (ii) peptides with non-conservative substitutions in positions strictly required for TCR recognition[14,22] were removed (**Figure 1a**). We searched the NCBI protein database for microbial peptide sequences with high homology to the nine amino-acid core of the immunodominant, deamidated CeD epitopes (DQ2.5-glia-α1a (PFPQPELPY), DQ2.5-glia-ω1 (PFPQPEQPF) and DQ2.5-glia-α2 (PQPELPYPQ), and further limited search parameters to microbes commonly associated with infections, the human microbiota, or reported to be in association with CeD. After removing duplicates and homologous sequences with gaps or insertions, we further selected sequences that conformed with the consensus peptide fine specificity of HLA-DQ2.5 restricted T cells from CeD patients (**Figure 1a**) [14,22]. Among the candidate peptides we noted a set of homologous, overlapping peptides from the *Pseudomonas fluorescens* protein succinylglutamate desuccinylase (PFSGDS), that mimicked the overlapping gliadin epitopes DQ2.5-glia-α1a and DQ2.5-glia-α2. We selected 23 minimal candidate peptides for further study (**Figure 1b**). These included variants of the overlapping mimic peptides from different *P. fluorescens* strains (DQ2.5-*P.fluor-α1a, -α2a* and DQ2.5-*P.fluor-α1b -α2b*) and four non-bacterial peptides that were identified in archaea and yeast. Accordingly, the structurally-guided database search identified candidate mimics of a number of HLA-DQ2.5-restricted gliadin epitopes.

Cross-reactivity between bacterial mimics and gliadin epitopes

To determine the extent of T cell cross-reactivity between each candidate mimic and gliadin epitope, we tested T cell responses using SKW3 cells that had been transduced with TCRs previously characterised from CeD patients. We used the TCRs LS2.8/3.15, N12, L6, JR5.1 and S16, which had established specificities for the CeD epitopes, DQ2.5-glia-α1a, DQ2.5-glia-ω1 and DQ2.5-glia-α2[14]. The candidate peptides were tested in two groups, namely the DQ2.5-glia-α1a / DQ2.5-glia-ω1 mimics, and the DQ2.5-glia-α2 mimics. The DQ2.5-glia-α1a/ω1 mimics were tested against SKW3.LS2.8/3.15-TCR and SKW3.N12-TCR cell lines, which both recognise DQ2.5-glia-

α1a, and against SKW3.L6-TCR cell line that cross-reacts with DQ2.5-glia-α1a and DQ2.5-glia-ω1[22]. The DQ2.5-glia-α2 mimics were tested against SKW3.JR5.1-TCR and SKW3.S16-TCR cell lines, whose TCRs express the well-characterised, TRAV26-1/TRBV7-2 biased TCR usage with a conserved arginine within the CDR3β loop. Stimulation assays established that each T cell line was readily activated by the expected gliadin peptide at 32 μg/ml and responded to at least one of the 10- and 11-mer mimic peptides in the presence of HLA-DQ2^{+} antigen presenting cells **(Figure 2)**. To obtain a relative measure of potency, we tested T cell line responses in the presence of peptide concentrations ranging from 1 - 32 μg/ml for the DQ2.5-glia-α1a/ω1 and DQ2.5-glia-α2 mimics, respectively (**Supplementary Figure 1a and 1b**), and confirmed HLA-DQ restriction by testing T cell line responses in the presence anti-DR, anti-DP and anti-DQ blocking antibodies (**Supplementary Figure 2a)**.

The SKW3.LS2.8/3.15-TCR, SKW3.N12-TCR and SKW3.L6-TCR cell lines responded moderately to the canonical DQ2.5-glia-α1a epitope at higher concentrations, and only the DQ2.5-glia-α1a/DQ2.5-glia-ω1 cross-reactive SKW3.L6-TCR responded to DQ2.5-glia-ω1 (**Figure 2a, Supplementary Figure 1a)**. One of the mimic epitopes, DQ2.5-*P.fluor-α1a,* was recognised by all three cell lines and elicited markedly stronger responses in the SKW3.LS2.8/3.15-TCR and SKW3.N12-TCR cell lines than the canonical DQ2.5-glia-α1a (**Figure 2a**). The T cell response to the related epitope DQ2.5-*P.fluor-α1b*, from a variant *P. fluorescens* strain was similarly strong in the SKW3.N12-TCR cell line, but weaker in the SKW3.LS2.8/3.15-TCR cell line and absent in the SKW3.L6-TCR line. In addition, the SKW3.N12-TCR line responded to DQ2.5-*E.cloac-ω1a* from *Enterobacter cloacae* at higher concentrations, and the mimic DQ2.5-*A.baum-α1a* from *Acinetobacter baumanii* elicited weak responses in SKW3.L6-TCR and SKW3.N12-TCR (**Figure 2a, Supplementary Figure 1a)**.

The SKW3.JR5.1-TCR and SKW3.S16-TCR cell lines were strongly activated by DQ2.5-glia-α2, but showed significant differences in their responses to the bacterial mimic peptides. Namely, SKW3.JR5.1-TCR was strongly activated by the mimics DQ2.5-*P.aerug-α2a*, DQ2.5-*P.fluor-α2a*, and by the variant DQ2.5-*P.fluor-α2b*. In contrast, the SKW3.S16-TCR cell line strongly responded to DQ2.5-*B.copro-α2a*, but only marginally responded to DQ2.5-*P.fluor-α2a* and not to DQ2.5-*P.aerug-α2a* or DQ2.5-*P.fluor-α2b* (**Figure 2b, Supplementary Figure 1b**). The discordant specificity of JR5.1 and S16 may be related to how these TCRs differ in their degree of HLA-DQ2.5-glia-α2 engagement[14]. Thus, there are distinct patterns of TCR cross-reactivity between distinct gliadin epitopes and the bacterial mimics.

The parental bacterial protein is naturally processed and presented

An observation arising from previous stimulation experiments was that, when combined, the overlapping epitopes DQ2.5-glia-α1a and DQ2.5-glia-α2 elicited stronger responses in T cell lines. We therefore tested SKW3 T cell activation by longer versions of mimic epitopes (**Supplementary Figure 2**) and investigated the parental protein (PFSGDS) from which the overlapping epitopes DQ2.5-*P.fluor-α1a* and DQ2.5-*P.fluor-α2a* were derived. The PFSGDS protein was expressed in *E. coli* and purified by immobilized metal affinity chromatography followed by gel filtration (**Supplementary Figure 3**). The immunogenicity of the purified PFSGDS protein was subsequently tested in the SKW3.TCR cell line stimulation tests, and several of the lines responded although generally weaker when compared to the 10-mer peptides DQ2.5-*P.fluor-α1a* and DQ2.5-*P.fluor-α2a* (**Figure 2 and Supplementary Figure 1**). We reasoned that inefficient stimulation was either due to limited uptake of protein by the B lymphoblastoid cell line (BLCL) used as antigen-presenting cells in this assay (BLCL), or due to inadequate processing and presentation of the full-length protein by HLA-DQ2.5. To exclude the latter possibility, we incubated $0.5*10^9$ antigen presenting cells (HLA-DQ2.5$^+$ 9022 BLCL, or HLA-DQ8$^+$ 9033 BLCL) in the presence of 300 µg/ml of PFSGDS overnight and subsequently isolated cellular HLA-DQ via immunoprecipitation. The bound peptides were eluted and analysed via mass spectrometry as previously described [23]. Amongst the HLA-DQ2.5 peptides identified from the antigen fed 9022 BLCL, a single nested set of peptides covering the sequence of both DQ2.5-*P.fluor-α1a* and DQ2.5-*P.fluor-α2a* **(Figure 3a)** was detected. In contrast, peptides isolated from antigen –exposed 9033 BLCL expressing HLA-DQ8 produced no such match against the sequence of PFSGDS **(Figure 3a)**. To confirm the identity of DQ2.5-*P.fluor-α1a/2* derived peptides*, the* chromatographic retention and fragmentation spectra of the synthetic 15-mer peptide derived from this region of DQ2.5-*P.fluor-α1a/2a* was compared to the relevant experimentally-derived peptide (**Figure 3b, Supplementary Figure 2**). This analysis confirmed that the PFSGDS protein was processed and presented by HLA-DQ2.5$^+$ BLCL but not by HLA-DQ8$^+$ BLCL. Since the 15-mer DQ2.5-*P.fluor-α1a/2a* peptide includes the two predicted overlapping epitopes, we tested this peptide for recognition by each of the SKW3 T cell lines (**Supplementary Figure 2**). Notably, this peptide elicited strong signals in all SKW3.TCR cell lines tested, including S16, which only marginally responded to DQ2.5-*P.fluor-α2a* (**Supplementary Figure 1b**). This indicated that SKW3.S16-TCR cell line was sensitive to the overall length of the bacterial mimic peptide. Moreover, these results suggested that the PFSGDS protein is immunogenic for a broad range of T cell lines tested.

CeD patient-derived T cells proliferate in response to bacterial peptides

Next, to gauge the immunogenicity of the PFSGDS protein for CeD-associated T cells in a more physiological setting we determined T cell proliferation of CeD patient-derived T cell clones using HLA-DQ2.5^{+} PBMCs as antigen-presenting cells (**Figure 3c**). Overall, 4 of the 9 patient-derived T cell clones were restricted to HLA-DQ2.5-glia-α1a (25-5101, N10, N12 and L6), and 5 to HLA-DQ2.5-glia-α2 (S16, 25-5204, 136-009, M402, D1). Moreover, the T cell clones included 3 (N12, L6 and S16) that matched the TCRs of the SKW3 TCR cell lines used in the mimic peptide screen. We measured proliferation via ^{3}H-Thymidine incorporation in response to PFSGDS protein and two 13-mer peptides, *DQ2.5-P.fluor-α1a-13mer and DQ2.5-P.fluor-α2a-13mer*, each containing both 9-mer cores of the mimic epitopes DQ2.5-*P.fluor-α1a* and DQ2.5-*P.fluor-α2a* **(Supplementary Figure 2)**. With the exception of L6, each of the patient-derived T cell clones responded to one or both of the 13-mer mimic peptides with a response comparable to the canonical gliadin peptide, and the majority of TCCs (except L6 and 136-009) responded vigorously to the PFSGDS protein. The T cell clones 25-5101 and N10 failed to respond to DQ2.5-*P.fluor-α2a* suggesting that both T cell clones required the N-terminally extended core peptide for recognition. We subsequently repeated the experiment for HLA-DQ2.5-glia-α1a restricted TCCs N10, N12, K5103 and 25-5204 using different concentrations of PFSGDS and peptide (**Figure 3d**) and observed that each clone was highly sensitive to the bacterial protein, with a half-maximal response at significantly lower concentrations than for the synthetic peptides. No stimulation was observed in the absence of allogeneic PBMC antigen presenting cells (APC), thus excluding stimulation from LPS/endotoxin activity in the PFSGDS preparation (data not shown). This data suggests that a sub-population of CeD associated T cells are potently activated by a bacterial protein contained in particular *P. fluorescens* strains.

Gluten challenge triggers response to bacterial mimic reactive T cells in CeD patients

During a gluten free diet, the frequency of circulating, gliadin reactive T cells is typically very low in CeD patients' blood and increases within days of gluten consumption. Oral gluten challenge therefore provides a readout of a circulating, polyclonal memory-recall T cell response to gluten that is disease specific [24]. Hence, we used IFN-γ ELISpot assays to measure the reactivity of gluten-specific T cells in the blood of 5 CeD patients to mimic peptides and PFSGDS before (day 0) and on day 6 after a 3-day wheat gluten challenge to the mimic peptides and PFSGDS protein. ELISpot assays were set up with a concentration range of 10-200 μg/ml of the mimic peptides and 100-1000 μg/ml PFSGDS **(Figure 4)**. On day 6 after gluten challenge, the two gliadin peptides encompassing overlapping epitopes DQ2.5-glia-α1a/α2 and DQ2.5-glia-ω1/ω2 (50 μg/ml), used as positive controls, led to significant responses in all 5 patients, whereas significant responses to 50 μg/ml of the mimic peptides DQ2.5-*P.fluor-α1/α2a-15mer*, DQ2.5-*P.fluor-α1/α2b-15mer*, DQ2.5-*P.aerug-*

α2a-13mer and DQ2.5-*B.copro-α2.2-13mer* showed a somewhat lower penetrance of 2/5, 1/5, 2/5 and 2/5 patients (**Figure 4a-d)**, respectively, and 100 μg/ml PFSGDS induced significant responses in the same patients (#670 and #496) as the peptide DQ2.5-*P.fluor-α1a-15mer* **(Figure 4e),** which is encompassed by PFSGDS. Notably, the patient (#670) showed the strongest ELISpot responses to the canonical gliadin peptides and to all bacterial peptides. As with the canonical gliadin peptides, the T cell responses to the bacterial mimics and PFSGDS only reached significant levels following gluten consumption, consistent with a gluten-specific recall response **(Figure 4f)**. Thus, our data shows that T cell responses to mimic peptides in patients' blood are induced by gluten consumption and partially overlap with T cell responses to established gliadin epitopes.

Differential TCR affinity towards deamidated gliadin and mimic epitopes

LS2.8/3.15 and N12 TCR transduced T cell lines were more strongly activated by the bacterial mimic peptide DQ2.5-*P.fluor-α1a* than by DQ2.5-glia-α1 (**Figure 2a, Supplementary Figure 1a**), and the T cell lines JR5.1 and S16 differentially responded to the DQ2.5-glia-α2 mimics DQ2.5-P.aerug-α2a and DQ2.5-B.copro-α2a (**Figure 2b, Supplementary Figure 1b**). To investigate the basis for the differential activation by the bacterial mimic peptides, we expressed, refolded and purified (i) the TCRs LS2.8/3.15, L6, JR5.1 and S16, and (ii) the peptide-HLA-DQ2 complexes with the mimic peptides DQ2.5-*P.fluor-α1a,* DQ2.5-*P.fluor-α1b* and DQ2.5-P.aerug-α2a, and with the corresponding DQ2.5-glia-α1 and DQ2.5-glia-α2 epitopes. We then performed surface plasmon resonance (SPR) measurements to determine the affinities of the purified TCRs for these different HLA-DQ2.5 complexes (**Figure 5**). The SPR measurements reflected the observations from the peptide screening experiments, and indicated that the observed differences in T cell line activation assays directly correlated with differences in binding affinities. The TCR LS2.8/3.15 bound to HLA-DQ2.5-*P.fluor-α1a* with a significantly higher affinity (39.6 μM) than DQ2.5-glia-α1 and HLA-DQ2.5-*P.fluor-α1b* (91.5 μM and >200 μM, respectively), while the L6 TCR had a clear preference for HLA-DQ2.5-glia-α1 (21.4 μM) and only bound weakly to HLA-DQ2.5-*P.fluor-α1a* (>200 μM), and showed no detectable binding to HLA-DQ2.5-*P.fluor-α1b* (**Figure 5a**). The results with the HLA-DQ2.5-glia-α2 restricted TCRs JR5.1 and S16 similarly confirmed observations from the T cell line stimulation assays, namely, JR5.1 bound to HLA-DQ2.5-glia-α2 and to HLA-DQ2.5-P.aerug-α2a with similar affinities (83.7 μM and 132 μM, respectively), whereas the S16 TCR only bound to HLA-DQ2.5-glia-α2 (13.5 μM), but not to HLA-DQ2.5-P.aerug-α2a (**Figure 5b**). Accordingly, the affinity values of CeD TCRs towards bacterial mimics corresponded to that of their stimulatory capacity.

Bacterial ligand structural mimic of DQ2.5-glia-α1 epitope

To establish how a bacterial mimic of the DQ2.5-glia-α1 epitope was presented by HLA-DQ2.5, we determined the crystal structure of the binary HLA-DQ2.5-*P.fluor-α1a* complex to 1.9 Å resolution (**Table 1)** and compared it to the structure of the HLA-DQ2.5-glia-α1 complex. The HLA-DQ2.5-*P.fluor-α1a* structure aligned very well with the HLA-DQ2.5-glia-α1 crystal structure, with a root mean squared deviation (r.m.s.d) = 0.3Å for all Cα atoms of the peptide binding cleft, and both peptides were bound in a very similar conformation (**Figure 6a**). With the sequence variations between the two peptides located in the p2 and p4 positions, the only notable difference in the peptide was evident at the solvent exposed p2 position, where the p2-Met residue DQ2.5-*P.fluor-α1a* was distinct from p2-Phe DQ2.5-glia-α1. Nevertheless, the p4-Met residue in DQ2.5-*P.fluor-α1a* was, despite its different chemical properties, almost perfectly aligned with p4-Gln in DQ2.5-glia-α1 (**Figure 6a**). Accordingly, the DQ2.5-*P.fluor-α1a* epitope was a close structural mimic of the DQ2.5-glia-α1 epitope.

Structure of the TCR-HLA-DQ2.5-*P.fluor-α1a* complex

To establish the molecular basis for how a HLA-DQ2.5-glia-α1 reactive TCR could recognise the bacterial mimic DQ2.5-*P.fluor-α1a*, we determined the crystal structure of the LS2.8/3.15 TCR-HLA-DQ2.5-*P.fluor-α1a* complex to 2.8 Å resolution (**Table 1, Figure 6b-f**). The TCR docked on HLA-DQ2.5-*P.fluor-α1a* in a conventional orientation (**Figure 6b**), with an overall buried surface area (BSA) of 2200 $Å^2$. The footprint of the LS2.8/315 TCR on HLA-DQ2.5-*P.fluor-α1a* was dominated by the CDR3β loop, which contributed 37% of the BSA, and the remainder of the interface was made up of smaller contributions by the remaining CDR loops and both α- and β-framework residues (**Figure 6b**). The number of interactions the LS2.8/3.15 TCR made with the pHLA were markedly skewed towards the HLA β-chain, which was reflected by the large BSA contribution of 54.1% made by the HLA β-chain compared to 28.8% and 17.7% by the HLA α-chain and peptide, respectively. Despite distinct CDR3 sequences and TRAV/TRBV usage, the LS2.8/315 TCR revealed some degree of resemblance to that of the S2 TCR-HLA-DQ2.5-glia-α1 complex[14] in that the S2 TCR had an overall similar layout of CDR loops and covered a comparable area on the pHLA with a nearly identical BSA of 2200 $Å^2$ and a large BSA contribution (54.8%) of the HLA β-chain.

Both CDR3α and CDR3β loops of LS2.8/3.15 traversed the peptide binding cleft and interacted with the HLA α- and β-chains (**Figure 6c**). Whilst the comparatively short CDR3α loop made limited contacts to Phe58α and to Arg77β of HLA-DQ2.5, the longer CDR3β loop adopted a brace-like conformation atop the central portion of the antigen-binding cleft. The interface with the HLA α-chain was formed by Gln111β at the tip of CDR3β loop, which was wedged between the

sidechains of Phe58α, Thr61α and Asn62α of the HLA α-chain helix, and formed H-bonds to Asn62α and the backbone of Phe58α (**Figure 6d**). These interactions were further enhanced by the adjacent β-framework reside Arg66β, which formed vdw interactions with HLA residues Gln57α, Phe58α and Thr61α (**Figure 6e**). On the opposing side of the peptide, the CDR3β loop formed an extensive vdw interface with the ridge of the HLA β-chain, involving residues Tyr60β, Gln64β, Asp66β, Ile67β and Arg70β (**Figure 6d**). Moreover, this interface was further extended to Glu69β, Ala73β and Arg77β through vdw interactions with CDR1α and CDR2α loops and a salt bridge with the TCR α-chain framework residue Lys66 (**Figure 6f**).

The contacts between the LS2.8/315 TCR and the DQ2.5-*P.fluor-α1a* peptide were exclusively mediated through vdw interactions with hydrophobic peptide sidechains in p3-Pro, p5-Pro, p7-Leu and p8-Pro (**Figure 6c**). Here, Ser110α from the CDR3α loop interacted with p3-Pro, and the CDR3β residues Glu109β, Gln111β and Ala113β formed a larger contact area involving p5-Pro, p7-Leu and p8-Pro. Moreover, p8-Pro interacted with Lys37β and Tyr57β from the CDR1β and CDR2β loops, respectively. Accordingly, the LS2.8/315 TCR did not contact the P2 and P4 positions of the bacterial DQ2.5-*P.fluor-α1a* mimic that differed from the DQ2.5-glia-α1 epitope. Thus, cross-reactivity was associated with the TCR seeing the similarities between the bacterial and gliadin epitopes.

Molecular mimicry drives TCR cross-reactivity.

To establish the molecular basis for TCR recognition of the DQ2.5-glia-α2 bacterial mimic, we determined the structure of the JR5.1 TCR-HLA-DQ2.5-*P.aeru-α2a* complex at 2.8Å resolution (**Table 1, Figure 7**) and compared this to the JR5.1 TCR-HLA-DQ2.5-glia-α2 complex structure[14] **(Figure 7)**. The overall structure of the JR5.1 HLA-DQ2.5-*P.aeru-α2a* complex was very similar to that of the corresponding HLA-DQ2.5-glia-α2 complex, with only a slight tilt (2-3º) in the long axis of the TCR towards the peptide N-terminus (**Figure 7a**) and an overlay of the peptides DQ2.5-*P.aeru-α2a* and HLA-DQ2.5-glia-α2 only revealed very minor positional variations (Cα r.m.s.d. = 0.17Å) (**Figure 7b**). Accordingly, the footprints of the JR5.1 TCR on HLA-DQ2.5-*P.aeru-α2a* (**Figure 7c**) and HLA-DQ2.5-glia-α2 (**Figure 7d**), and) show the interactions with the DQ2.5-*P.aeru-α2a* peptide (**Figure 7e**) and the DQ2.5-glia-α2 peptide (**Figure 7f**) were very similar. The interactions between the HLA and the JR5.1 TCR CDR3 loops (**Supplementary Figure 4a**), and the germline-encoded residues (**Supplementary Figure 4b and 4c**) were essentially conserved between both complexes. Being slightly tilted towards the peptide N-terminus in the HLA-DQ2.5-*P.aeru-α2a* complex, the JR5.1 TCR formed additional interactions with the peptide N-terminal

side and lost some interactions on the C-terminal side of the peptide. Additional interactions in the HLA-DQ2.5-*P.aeru-α2a* complex were a vdw-contact between Asn36α and p2-Gln of the peptide (**Figure 7e**) and two H-bonds between the backbone of the TRBV-framework residue Asp67β and the HLA α-chain residues Lys39α and Gln57α (**Supplementary Figure 4b**). Interactions lost in the JR5.1 TCR-HLA-DQ2.5-*P.aeru-α2a* complex were located near the C-terminus of the peptide, where CDR3β Phe108 was lifted upwards by 1Å and thereby lost some of the vdw interactions with the HLA β-chain (**Supplementary Figure 4a**). These subtle differences in binding were consistent with the angular shift in TCR docking. Accordingly, molecular mimicry drives the TCR cross reactivity across the gliadin and bacterial epitope.

Discussion

The *HLA II locus* is strongly associated with a number of T cell mediated autoimmune disorders, where various mechanisms have been implicated [2]. For example, in CeD, deamidation of gluten peptides enables binding to HLA-DQ2/DQ8 molecules, thereby facilitating an acquired, aberrant $CD4^+$ T cell response. Nevertheless, the lack of penetrance of a given *HLA allele* to cause disease indicates that other factors, including genetic and environmental, play a role in precipitating the disease. The nature of the triggers and drivers relating to HLA-associated IMIDs remains unclear.

Population studies have implicated a range of environmental factors associated with CeD risk. These include infections, particularly gastrointestinal infections, gluten feeding practices, medications and perinatal factors such as season of birth [19]. An emerging concept is that these environmental factors contribute to the disruption of oral tolerance [25]. Specific viral infections linked to CeD include rotavirus [26], adenovirus [20,21], reovirus [27] and enterovirus [28]. However, as the majority of CeD patients in these studies did not have evidence of prior viral exposure, other triggers for CeD development are likely. A range of studies now support a secondary role for opportunistic bacterial pathogens in CeD [18,29-37]. Proteases from commensal gut bacteria are capable of degrading gluten proteins and releasing immunogenic peptide fragments more amenable to absorption, and can directly exacerbate gluten immunopathology in a HLA-DQ8 mouse model [36,37]. *P fluorescens* is a minor component of the gut microbiota in humans, a rare cause of blood-borne infection in humans [38-40], and IgA antibodies to *P.fluorescens* are associated with Crohn's disease [41,42]. In CeD, positive anti-*P.fluorescens* serology was observed in 86% of patients versus 31% of healthy controls [29]. Interestingly, sero-reactivity to microbial markers such as *Pseudomonas fluorescens*-associated sequence I2 is seen in early CeD [18] as well as CeD poorly responsive to a gluten-free diet [34]. This finding raises the possibility that immunoreactivity to microbial antigens

may play a role in CeD development or perpetuation of mucosal inflammation in patients avoiding gluten.

There are several, potentially synergistic, mechanisms by which microbes are postulated to contribute to the loss of T cell tolerance to gluten. Molecular mimicry, whereby T cells target microbial antigens that mimic particular (self-)antigens associated with autoimmune disease, provides a model for the initial phase of (self-)antigen sensitisation via cross-reactive T cells. This mechanism was proposed in CeD based on sequence homology between adenovirus and gliadin but this has never been substantiated [20,21]. Our findings provide evidence that molecular mimicry to an exogenous bacterial antigen may be a plausible primary mechanism contributing to the abrupt onset of disease in HLA-DQ2^{+} individuals, usually in infancy, by inducing cross-reactive gluten-specific CD4^{+} T cells. Colonisation of the intestinal mucosa by *P fluorescens* may also be a potential explanation for persistent mucosal injury in CeD patients strictly avoiding dietary gluten. We have identified peptide mimics expressed in a range of related bacterial species. For example, DQ2.5-*P.aeru-α2a* is present in several *Pseudomonas* and *Bordetella* species. Our findings support the data that *P. aeruginosa* can modulate CeD inflammation and provides another mechanism distinct to the effects of *P. aeruginosa* elastase [36,37]. Accordingly, our data demonstrates that the prerequisites exist for involvement of bacterial mimic peptides in the triggering of gluten-specific CD4^{+} T cells from CeD patients, but more data is required in order to causally link specific bacteria to CeD aetiology.

TCR recognition is highly degenerate and the sequence differences between canonical peptides and molecular mimic peptides can, in principle, be quite substantial. Nevertheless, we used an approach of a combined sequence homology search of microbes commonly associated with infections, the human microbiota, or reported to be in association with CeD with structural and functional data on CeD associated T cells. We identified peptides from common commensal and pathogenic bacteria that cross-reacted with CeD patient-derived T cells restricted to the immunodominant gliadin epitopes DQ2.5-glia-α1 and DQ2.5-glia-α2. Our structural and functional data identified that molecular mimicry was the underlying molecular mechanism for T cell cross-reactivity. To date, there are only a few examples of candidate mimic epitopes in IMID that have been identified [43 44 45]. Our study exemplifies how mimic epitopes can be identified using both homology and reactivity of disease associated T cells, and demonstrates that common and abundant bacteria do express highly active mimic antigens that closely resemble two immunodominant and deamidated epitopes targeted by pathogenic T cells in CeD.

Thus, our study highlights the possibility that CeD results from the triggering of a host T cell response to bacteria that cross-reacts with gluten epitopes. Once initiated this may be followed by gluten driven expansion of the cross-reactive T cell receptor repertoire and epitope spreading. This may explain why only a minority of the gluten-specific T cell response induced by oral gluten challenge responds to the bacterial mimic epitopes. Finally, the high frequency of mimic peptides in our sample set of candidate peptides suggests that bacteria likely contain a pool of mimic antigens that fit into the context of other autoimmune diseases. Ultimately, such knowledge may be used to prevent IMIDs in genetically predisposed individuals.

Acknowledgements

We thank the staff at the Australian Synchrotron for assistance with data collection, the staff at the Monash Macromolecular crystallization facility. We thank the CeD volunteers who participated in this study. This work was supported by a program grant from the National Health and Medical Research Council of Australia (NHMRC) and the Australian Research Council (ARC) (CE140100011). FK is supported by the collaboration project TIMID (LSHM18057-SGF) financed by the PPP allowance made available by Top Sector Life Sciences & Health to Samenwerkende Gezondheidsfondsen (SGF) to stimulate public-private partnerships and co-financing by health foundations that are part of the SGF. AWP is supported by an NHMRC Principal Research Fellowship; JR is supported by an Australian ARC Laureate Fellowship.

Competing interests Declaration

RPA is an employee of ImmusanT, Inc., a company developing a celiac disease immunotherapy. RPA and JAT-D are inventors of patents, owned or licensed by ImmusanT, Inc., relating to the diagnostic application of gluten challenge, and utilisation of gluten-derived T cell epitopes for use in therapeutics.

Figure legends

Figure. 1: Identification of microbial mimic peptides for CeD epitopes using TCR structural and functional data. (a) Location of TCR – peptide interactions in published TCR-pHLA structures and TCR sensitivity to individual peptide substitutions in HLA-DQ2.5-glia-α1, HLA-DQ2.5-glia-ω1 and HLA-DQ2.5-glia-α2. Peptide surface representations from TCR-pHLA structures are coloured according to contacting CDR loop (red: CDR1α, cyan: CDR3α, orange: CDR1β, purple: CDR2β, blue: CDR3β). TCR sensitivity to Ala mutation is expressed (blue bars) as fraction of TCRs that loose >75% function upon Ala substitution in p1-p9 of the peptide. (b) Selected mimic peptides with negative charge at canonical deamidation site and substitution score based on TCR sensitivity.

Figure. 2: Bacterial mimic peptides and parent protein potently stimulate CeD patient derived T cells. Screening of bacterial mimic peptides for activation of TCR transduced of SKW3 TCCs. CD69 and CD3 expression levels of (a) DQ2.5-glia-α1 restricted TCCs and (b) DQ2.5-glia-α2 restricted TCCs were measured after 16h incubation the presence HLA-DQ2^{+} BLCLs and 32 μg/ml peptide or 500 μg/ml PFSGDS protein. The percentage of CD69^{high} and CD3^{low} TCCs (mean ± standard deviation (SD) from 2 independent experiments with 2 replicates) was used as a relative indicator of T cell activation. Responses were considered positive where the mean %CD69^{high} signal exceeded background (DMSO) plus 2 times SD, and significant, where CD69^{high} was more than 3 SD above background.

Figure. 3: HLA-DQ2^{+} antigen presenting cells process PFSGDS and present antigenic mimic peptide, and patient derived T cells proliferate in response to bacterial mimic peptides or PFSGDS protein. Peptides derived from succinylglutamate desuccinylase were identified following antigen feeding of HLA-DQ2.5+ or HLA-DQ8+ cells and subsequent elution of HLA-DQ-bound peptides and sequencing by LC-MS/MS. (a) Position 261-299 of succinylglutamate desuccinylase reveals disparity of nested peptide presentation between DQ2- and DQ8-expressing cells. Region in bold indicates peptide with highest confidence identification score. (b) b-, y- and internal-fragment ion assignments for peptide GEPmPmPELPYPATP (lowercase m denotes oxidised methionine) eluted from 9022 (DQ2+) cells. (c, d) Proliferation of CeD patient derived T cell lines induced by mimic peptides and PFSGDS protein. (c) Proliferation of DQ2.5-glia-α1 restricted and DQ2.5-glia-α2 restricted T cell lines (left and right of dashed line, respectively) (d) Proliferation of patient derived DQ2.5-glia-α1 restricted T cell lines induced by dilutions of mimic peptides and PFSGDS protein. In total 19 independent T cell clones were tested against gliadin and

mimic peptide epitopes in 18 experiments. One T cell clone was tested 5 times, 4 were tested 4 times, 9 were tested 3 times, 3 were tested twice, 2 were tested once. In 9 independent experiments the response to the bacterial PFSGDS protein was determined as well. In 5 of those experiments the T cell clones were tested against a concentration range of the PFSGDS protein. Representative results are shown.

Figure. 4: T cell cross-reactivity following gluten challenge. IFN-γ ELISpot assays show significant T cell reactivity to mimic peptides and PFSGDS protein in PBMCs from five CeD patients after wheat gluten challenge. T cell activation (spot-forming units; SFU/10^6 PBMCs) was measured in the presence of varying concentrations of the mimic peptides: (a) DQ2.5-P.fluor-α1/2a-15mer, (b) DQ2.5-P.fluor-α1/2b-15mer, (c) DQ2.5-P.aerug-α2a-13mer, (d) DQ2.5-B.copro-α2.2-13mer, and (e) PF SGDS protein; μg/mL concentrations shown on each graph. T cell responses against 50μg/mL control gliadin peptides (containing DQ2.5-glia-α1a/α2 and DQ2.5-glia-ω1/ω2) are shown on each graph as a comparison. Mean ± SD of duplicate wells are shown. * Responses above the cut-off are considered positive. (f) T cell responses against 100μg/mL bacterial peptide, 250μg/mL protein, or 50μg/mL control gliadin peptides in all patients prior to (day 0) and day 6 after gluten challenge in the same five CeD patients. Response cut-offs are depicted as dotted lines.

Figure. 5: Surface plasmon resonance affinity measurements. (a) Binding affinities of DQ2.5-glia-α1 restricted TCRs LS3.15 and L6 were determined for surface bound HLA-DQ2.5-glia-α1, HLA- DQ2.5-*P.fluor-α1a*, and HLA-DQ2.5-*P.fluor-α1b*. (b) Binding affinities of DQ2.5-glia-α2 restricted TCRs JR5.1 and S16 were determined for surface bound HLA-DQ2.5-glia-α2 and HLA-DQ2.5-*P.aerug-α2a.*

Figure. 6: Structural basis for the recognition of HLA-DQ2.5-*P.fluor-α1a* (a) Overlay of the crystal structures of the binary complexes HLA-DQ2.5-*P.fluor-α1a* (purple sticks) and HLA-DQ2.5-glia-α1 (grey sticks) shows good alignment for the peptides. b-f) Crystal structure of the LS2.8/3.15 TCR - HLA-DQ2.5-*P.fluor-α1a* ternary complex. Colours: CDR loops and corresponding TCR footprint contacts are coloured red, pink, cyan, orange, purple and blue for CDR1α, CDR2α, CDR3α, CDR1β, CDR2β, CDR3β, respectively, and framework residues are coloured green. The peptide and HLA α- and β-chains are coloured grey, light green, and light yellow, respectively. (b) TCR docking angle, footprint and BSA contributions. (c) Interactions between the LS2.8/3.15 TCR and the DQ2.5-*P.fluor-α1a* peptide. (d) Interactions between the CDR3 region of LS2.8/3.15 and the HLA. (e) Germline encoded interactions between LS2.8/3.15 and the HLA α-chain. (f) Germline encoded interactions between LS2.8/3.15 and the HLA β-chain.

Figure. 7: Molecular mimicry drives cross-recognition of bacterial epitope HLA-DQ2.5-*P.aerug-α2a*. Comparison of the ternary complexes between the JR5.1 TCR and HLA-DQ2.5 with bound peptides HLA-DQ2.5-*P.aerug-α2a* and HLA-DQ2.5-glia-α2. Colours: CDR loops and corresponding TCR footprint contacts are coloured red, pink, cyan, orange, purple and blue for CDR1α, CDR2α, CDR3α, CDR1β, CDR2β, CDR3β, respectively, and framework residues are coloured green. The peptide and HLA α- and β-chains are coloured grey, light green, and light yellow, respectively. (a) Overlay of the ternary complex structures of the JR5.1 TCR with HLA-DQ2.5-*P.aerug-α2a* and HLA-DQ2.5-glia-α2 shows a small tilt in the angle in which the JR5.1 TCR binds to HLA-DQ2.5-*P.aerug-α2a* and HLA-DQ2.5-glia-α2. (b) Overlay of the peptides HLA-DQ2.5-*P.aerug-α2a* (grey sticks) and HLA- DQ2.5-glia-α2 (beige sticks) in the respective ternary complexes. (c, d) footprint and docking angle and BSA contributions of the JR5.1 TCR binding to c) HLA-DQ2.5-*P.aerug-α2a* and d) HLA-DQ2.5-glia-α2. (e, f) Comparison of the interface between the JR5.1 TCR and the peptides (e) DQ2.5-*P.aerug-α2a* and (f) HLA-DQ2.5-glia-α2.

Methods

Protein expression and purification

HLA-DQ complexes. The extracellular domains of HLA-DQ2.5 (HLA-DQA1*5:01 and HLA-DQB1*02:01) with the canonical gliadin epitopes and the HLA-DQ8-CLIP complex were produced as previously described [13,14]. For the bacterial epitopes the peptides DQ2-*P.fluor-α1a*: APMPMPELPYP, DQ2-*P.fluor-α1b*: APMPLPDLPYP, and DQ2-*P.aeru-α2a*: AMVVQSELPYPE were covalently linked to the N-terminus of the HLA-DQ2.5 β-chain using the linker sequence GSGGSIEGRGGSG. HLA-DQ complexes expressed in the supernatant of baculovirus infected in Hi5 insect cells were concentrated and diafiltrated into 10 mM Tris, pH 8, and 500 mM NaCl using a Cogent M1 TFF system (Merck Millipore) and subsequently purified via immobilised metal affinity (Ni Sepharose 6 Fast Flow; GE Healthcare), size exclusion (Superdex 200; GE Healthcare) and anion-exchange (HitrapQ; GE Healthcare) chromatography. Prior to crystallisation experiments, the C-terminal domains of the constructs were removed by cleavage with Enterokinase (New England Biolabs) followed by a second round of anion-exchange chromatography.

T cell receptors. The extracellular domains of the TCR α and β chains with an engineered interchain disulfide bond were expressed in E coli and refolded and purified as described previously [13].

PFSGDS. *E. coli BL21-DE3* transformed with a pET30 vector containing the *E. coli* codon optimised DNA sequence of PFSGDS in frame with a C-terminal His6-Tag was grown in LB media at 37 º C to OD_{600} =0.7 and induced with 500µM IPTG for 4 h at 25º C. Harvested cells were resuspended in ice cold 20 mM Tris pH8.0, 300mM NaCl, 0.5 mM tris(2-carboxyethyl)phosphine (TCEP) and 0.2 mM PMSF, and lysed by sonication. The lysate was cleared by centrifugation for 30 min at 25000 g, supplemented with 20mM Imidazole and passed over Ni-Sepharose. To remove contaminants and bacterial lipids, bound PFSGDS was extensively washed using 8 column volumes (CV) of buffer A (20 mM Tris pH8.0, 300mM NaCl, 20mM Imidazole, 0.2 mM TCEP), followed by 40 CV of buffer B (20 mM Tris pH8.0, 300mM NaCl, 0.2 mM TCEP, 0.1% Triton X100) and again 8 CV of buffer A. After elution with 300 mM Imidazole in buffer A, the protein was further purified via size exclusion chromatography (Superdex 200; GE Healthcare) in PBS.

Surface plasmon resonance

Surface plasmon resonance measurements were performed on a BIACORE 3000 instrument essentially as described previously [14]. Briefly, purified HLA-DQ2 complexes and HLA-DQ8-clip as background control were biotinylated and immobilised using a Biotin Capture Kit (GE healthcare, Parramatta, Australia) to a surface loading of 900-1500 response units. Serial dilutions of purified TCRs in SPR buffer containing (20mM HEPES pH7.4, 150 mM NaCl, 2 mM EDTA, 0.005% surfactant P20) were passed over the chip at a flow rate of 10 µL/min for 90s. two Two independent experiments with two replicates were performed for each TCR and equilibrium dissociation constants were determined by fitting a single site binding model to the data.

Crystallisation, diffraction data collection and structure determination

Crystals were grown at 20ºC via the hanging drop vapor diffusion method using equal volumes of protein (7-10 mg/ml in 10mM Tris-HCl pH 8.0, 150 mM NaCl) and crystallisation solutions. HLA-*DQ2-P.fluor-α1a* was crystallised with 18% PEG3350 and 0.15M $NH_4H_2PO_4$; the ternary complex LS2.8/3.15 – HLA-DQ2.5P.fluor-α1a with 22% PEG3350, 0.2M Na-K Tartrate, 0.08 M MES pH6.5; and the the ternary complex JR5.1 – HLA-*DQ2-P.aerug-α2a* in 23%PEG3350, 0.01mM Na-Acetate, 0.1 M Tris pH8.0. Prior to data collection crystals of HLA-*DQ2-P.fluor-α1a* and the two ternary complexes were cryoprotected in reservoir solution supplemented with 18 % glycerol or 20% PEG 400, respectively, and frozen in liquid N_2. X-ray diffraction data was collected at the Australian Synchrotron using the MX1 beamline for HLA-*DQ2-P.fluor-α1a* and LS2.8/3.15 - HLA-*DQ2-P.fluor-α1a*, and at the MX2 beamline for JR5.1 - HLA-*DQ2-P.aeru-α2a*. Crystals were exposed at 100K using a single wavelength (0.953725 Å, 0.94640 Å and 0.95372 Å, respectively) and the diffraction data was processed using XDS [46] and merged using Aimless of the CCP4

package [47]. The structures were solved via molecular replacement in Phaser [48] using previously published coordinates as search models; HLA-DQ2 (PDB code 6MFG) [22], and TCRs T316 and JR5.1 (PDB codes: 4Z7W and 4OZH) [13,14] for the LS2.8/3.15 – HLA-DQ2.5P.fluor-α1a and JR5.1 – HLA-*DQ2-P.aerug-α2a* structures, respectively. Model building, refinement and validation was carried out using Coot [49] and the Phenix software package [50]. The geometries of the refined structure models of were validated in Phenix, which indicated that HLA-*DQ2-P.fluor-α1a,* LS2.8/3.15 – HLA-DQ2.5P.fluor-α1a JR5.1 – HLA-*DQ2-P.aerug-α2a* contained 0%, 0.12% and 0.069% of Ramachandran outliers, respectively. The structure factors and refined atomic models were deposited in the PDB databank (PDB codes 6U3M, 6U3N, 6U3O).

Cell lines

SKW3 cells and TCR transduced SKW3 cell lines were maintained in RPMI-1640 supplemented with 50 IU/ml penicillin, 50 µg/ml streptomycin, 2mM Glutamine, 1x non-essential aminoacids, 1 mM Pyruvate, 10mM HEPES and 10% FBS ($RF10^+$ media); EBV-transformed B-lymphoblastoid cell lines 9022 (COX, $DQ2^+$; DQA1*05:01, DQB1*02:01) and 9033 (BM14, $DQ8^+$; DQA1*03, DQB1*0302) were maintained in the same media containing 15% FBS ($RF15^+$ media); Prior to SKW3 T cell stimulation and antigen feeding experiments all cells were rested for 24 h in fresh RF^+ media supplemented with 10% FCS, and cell density was adjusted to $2*10^6$ cells/ml. The B cell hybridomas SPV-L3 (anti–HLA-DQ), L243 (anti–HLA-DR) and b7/21 (anti–HLA-DP) were grown in RPMI-1640 supplemented with 5% FBS. TCR transduced cell lines SKW3.LS2.8/3.15, SKW3.L6, SKW3.N12, SKW3.JR5.1 and SKW3.S16 were produced via retroviral transduction of the αβTCR–deficient T cell leukemia cell line SKW-3, as previously described [51].

T cell proliferation assays

Proliferation assays were performed in triplicate in 150 µl Iscove's modified Dulbecco's medium supplemented with glutamine (Gibco) and 10% human serum in 96-well flat-bottom plates. Briefly, antigen-presenting cells (APCs) were loaded with antigen for 2 h, after which 20,000 gluten-specific T cells were added. As APCs we used 100.000 irradiated HLA-DQ2.5–matched allogeneic PBMCs (3,000 rad). Synthetic peptides were used at a final concentration of 6 µg/ml. After 48 h at 37 °C, cultures were pulsed with 0.5 µCi of 3H-thymidine and harvested 18 h later. As positive controls 13-mer versions of the DQ2.5-glia-α1 (LQPFPQPELPYPQ) and DQ2.5-glia-α2 (PFPQPEPLYPQPQ) epitopes were used.

SKW3 T cell stimulation assays

T cell stimulation assays monitoring CD69 and CD3 expression of SKW3.TCR cells were conducted in 96-well round bottom tissue culture plates: Serial dilutions of peptides (10mg/ml in DMSO), PFSGDS protein (34 mg/mL in PBS) or DMSO were set up in 50 µl media at room temperature, followed by addition of 2* 10^5 antigen presenting cells (9022 or 9033 BLCLs) and 10^5 SKW3.TCR cells (LS2.8/3.15, L6, N12, JR5.1 or S16) to a final volume of 200 µl/well. For antibody controls, antigen presenting cells were supplemented with blocking antibodies SPV-L3, L243 or b7/21 (5µg/well) prior to addition to the wells. Cells were incubated for 16 h (37º C, 5% CO_2) and subsequently washed with cold PBS and stained with Zombie Aqua™ viability dye (BD Pharmingen) followed by phycoerythrin-conjugated mouse anti-human CD69 (FN50; BD Pharmingen) and allophycocyanin-conjugated mouse anti-human CD3 (OKT3; BD Pharmingen). Cells were subsequently fixed with 2% Formaldehyde and stored in the dark at 4º C for up to 36 h prior to analysis. To determine surface expression levels of CD3 and CD69, the cells were analysed on a LSRFortessa™ X-20 instrument (BD-Biosciences) (refer to **Supplementary Figure 5** for gating strategy). Data were analysed using FlowJo 7.6 (Tree Star, OR, USA) and plotted in Prism (GraphPad San Diego, CA).

Antigen feeding and mass spectrometry

300 5*10^8 antigen presenting cells (9022 and 9033 BLCLs) were incubated for 16 h at 37º C in the presence of 300 µg/mL purified PFSGDS and subsequently harvested by centrifugation at 350g. After washing thrice with ice cold PBS, cell pellets were frozen in liquid N2 and stored at -80º C. Peptides presented by HLA-DQ were isolated via immunoprecipitation and analysed via as spectrometry as follows: Cell pellets were ground using a Retsch mixer mill (MM400) and lysed in 0.5% (v/v) NP-40, 50 mM Tris pH 8.0, 150 mM NaCl, and protease inhibitor cocktail (Roche cOmplete Protease Inhibitor) and HLA-DQ-peptide complexes isolated by immunoaffinity purification using protein A-crosslinked anti-DQ antibody (10mg per sample, clone SPV-L3). HLA-peptide complexes were dissociated by addition of 10% (v/v) acetic acid and peptides separated from heavy chain by RP-HPLC fraction on an Äkta Ettan (GE Healthcare) system using a Chromolith SpeedROD (RPC18 end-capped, 100 × 4.6-mm) column (Merck), over an increasing gradient mixture of buffer A (0.1% v/v trifluoroacetic acid (TFA) in water) and buffer B (80% v/v acetonitrile, 0.1% v/v TFA in water). Fractions were pooled using a concatenating scheme of every seventh fraction, dried down using a centrifugal concentrator (Labconco) and resuspended in 20 µL of mass spectrometery buffer A (0.1% v/v formic acid in water). Samples were analysed on a TripleTOF® 6600 (SCIEX) mass spectrometer, equipped to an on-line Eksigent Ekspert nanoLC 415 system (SCIEX) using a trap column (ChromXP C18, 3 µm 120 Å, 350 µm × 0.5 mm (SCIEX)) maintained at an isocratic flow of buffer A (2% v/v acetonitrile, 0.1% v/v formic acid in

water) at 5 µL/min for 10 min. Peptides were separated across an analytical column (ChromXP C18, 3 µm 120 Å, 75 µm × 15 cm (SCIEX)) by increasing linear concentrations of buffer B (80% v/v acetonitrile, 0.1% v/v formic acid in water) at a flow rate of 300 nL/min for 75 min. Up to 20 MS/MS spectra were acquired per cycle using an information dependent acquisition strategy with accumulation times of 200 ms and 150 ms for MS1 and MS2, respectively. MS1 scan range was set to 300-1800 m/z and MS2 set to 80-2000 m/z. To prevent multiple sequencing of the same peptide, MS1 masses were excluded for sequencing after two occurrences for 30 seconds. Data were analysed using PEAKS Studio v8.5 (Bioinformatics Solutions Inc) with the following settings: parent mass error tolerance of 50 ppm; fragment mass error tolerance of 0.1 Da; no enzyme cleavage; variable modifications of deamidation (NQ), phosphorylation (STY), oxidation (M). Data were searched against either the human proteome (Uniprot, November 2018) appended with the sequence of PFSGDS, or against PFSGDS alone.

Ex vivo stimulation of T cells from CeD patients after gluten challenge

Participants and gluten challenge

The study was approved by the Human Research Ethics Committees of the Walter and Eliza Hall Institute (no. 03/04) and the Royal Melbourne Hospital (no. 2003.009). CeD participants were adults aged between 18-70 years, were diagnosed according to ESPGHAN criteria[1], and following a gluten-free diet for at least 3 months prior to recruitment (**Supplementary table 1**). Whole blood samples were collected prior to (day 0, D0) and 6 days (D6) following 3-day oral wheat gluten challenge consisting of 4 slices of Bakers Delight white bread block loaf cut to toasting size thickness each day for 3 days. Symptoms were recorded on days 1-6 using the CeDPRO[2]. Baseline D0 blood was sent for serological screening for tissue transglutaminase-IgA and deamidated gliadin peptide-IgG levels at Melbourne Pathology.

IFN-γ ELISpot Assay

Peripheral blood mononuclear cells (PBMCs) from heparinized CeD patient whole blood were isolated using Ficoll-Paque Plus density-gradient centrifugation (GE Healthcare, Buckinghamshire, UK) within 4 hours of collection. Overnight IFN-γ ELISpot assays (Mabtech) were performed following modified manufacturer's instructions. ELISpot plates (Millipore Cat. #MSIPS4510) were coated with anti-human IFN-γ mAb (1- D1K;10µg/mL) overnight at 4℃. Plates were washed in PBS and blocked with 10% FCS in RPMI1640. Peptides were tested at a final concentration of 10-200 µg/mL and protein at 0.1-1 mg/mL added in a 25µL volume. PBMC were resuspended in complete medium: composed of RPMI 1640, 1x GlutaMAX, 100 %M NEAA from Gibco Thermo Fisher Scientific, 50 %M 2-ME (Sigma), 10% pooled human serum (Australian Red Cross Blood

691 Service), and 3-5×10^5 PBMC in 100μL were added per well. Control antigens tested included alpha
692 and omega wheat gliadin peptides containing the overlapping epitopes DQ2.5-glia-α1a/α2 and
693 DQ2.5-glia-ω1/ω2 (Figure 1a), 5 μg/mL Tetanus toxoid (Enzo Life Sciences), and 2.5 μg/mL
694 phytohemagglutinin-L (PHA-L; Sigma-Aldrich, St. Louis, MO). The following bacterial peptides
695 were screened: DQ2.5-P.fluor- α1/2a-15mer, DQ2.5-P.fluor-α1/2b-15mer, DQ2.5-P.aerug-α2a-
696 13mer and DQ2.5-B.copro-α2.2- 13mer (Figure 1b), as well as the PFSGDS protein. Following
697 overnight incubation, plates were washed with PBS/0.05% Tween 20 and PBS, and developed with
698 Mabtech anti-human IFN-γ secondary antibody (7-B6-1-Biotin; 1 μg/mL), Streptavidin-ALP (Cat.
699 #3420-2A; 1:1000 in 0.5% FCS/PBS), and 0.45 μM filtered Mabtech BCIP/NBT substrate (Cat.
700 #3650; 1:2 in distilled water). Spot-forming units (SFU) were counted using an automated ELISpot
701 reader (Autoimmun Diagnostika; Strassberg, Germany). Raw SFU were adjusted to SFU/10^6
702 PBMC to normalise between individuals. A cut-off of 20 SFU/10^6 PBMC was used to determine
703 positive responses, based on 5xSD calculated from the average of the 'No antigen' negative
704 controls from all patients. Mean of replicate wells and standard deviation measurements were
705 calculated and graphed using GraphPad prism version 7.0 (GraphPad, San Diego, CA).
706

Table 1: Data collection and refinement statistics

	HLA-DQ2.5-*P.fluor-α1a*	TCR LS2.8/3.15 - HLA-DQ2.5-*P.fluor-α1a*	TCR JR5.1 - HLA-DQ2.5-*P.aeru-α2a*
Data collection			
Space group	P 21 21 21	C 2 2 21	P 1 21 1
Cell dimensions			
a, *b*, *c* (Å)	94.935 96.276 105.74	59.98 239.47 147.46	69.057 157.65 106.1
α, β, γ (°)	90 90 90	90 90 90	90 96.53 90
Resolution (Å)	47.47 - 1.9 (1.968 - 1.9)	46.48 - 2.8 (2.9 - 2.8)	47.03 - 2.743 (2.841 - 2.743)
R_{pim}	0.05776 (0.4427)	0.02771 (0.4184)	0.04437 (0.3744)
I / □I	25.38 (2.16)	17.05 (1.85)	13.40 (2.00)
Completeness (%)	99.62 (96.94)	99.91 (99.92)	99.26 (98.60)
Redundancy	7.4 (7.5)	2.0 (2.0)	3.6 (3.4)
Refinement			
Resolution (Å)	47.47 - 1.9	46.48 - 2.8	47.03 - 2.743
No. reflections	76853 (7564)	26697 (2609)	58558 (5789)
R_{work} / R_{free}	0.1911 (0.2888) /0.2213 (0.3103)	0.2140 (0.3416) /0.2702 (0.3830)	0.2056 (0.3354) /0.2423 (0.3655)
No. atoms			
Protein	6008	6463	12766
Ligand/ion	99	14	56
Water	479	1	66
B-factors			
Protein	47.28	94.45	84.74
Ligand/ion	55.83	140.17	94.42
Water	45.65	60.47	56.02
R.m.s. deviations			
Bond lengths (Å)	0.007	0.004	0.002
Bond angles (°)	0.87	0.98	0.55

*Values in parentheses are for highest-resolution shell.

Supplementary data

Supplementary Figure Legends

Supplementary Figure. 1: Concentration dependent stimulation of SKW3 T cell clones with mimic peptides. Screening of bacterial mimic peptides for activation of TCR transduced of SKW3 TCC using different concentrations of peptide (0-32 μg/ml), or PFSGDS protein (0-500 μg/ml). CD69 and CD3 expression levels of (a) DQ2.5-glia-α1 restricted TCCs and (b) DQ2.5-glia-α2 restricted TCCs were measured after 16 h incubation in the presence of the indicated antigen and HLA-DQ2.5^{+} BLCLs. Data shown is percentage of cells with CD69^{high} or CD3^{low} (mean +/- SD calculated as from two independent experiments with two replicates). Responses were considered positive where CD69^{high} exceeded averaged background plus 2 times SD (<DMSO> + 2 SD), and significant, where CD69^{high} was more than 3 SD above averaged background.

Supplementary Figure. 2: Stimulation of SKW3 T cell clones with longer mimic peptides. (a) Screening of long bacterial mimic peptides (32 μg/ml) and PFSGDS (500 μg/ml) protein for activation CD69 of HLA-DQ2.5-glia-α1a and HLA-DQ2.5-glia-α2 restricted TCR transduced SKW3 TCC and specific blocking of stimulation by anti-HLA-DQ, but not anti-HLA-DR-, or anti-HLA-DP antibodies. CD69 expression levels of TCR transduced SKW3 TCC were measured in response to peptides or PFSGDS with HLA-DQ8^{+} or HLA-DQ2^{+} antigen presenting cells in the presence of HLA blocking antibodies. Data shown is percentage of cells with CD69^{high} averaged from two independent experiments with two technical replicates. Responses were considered positive where CD69^{high} exceeded averaged background plus 2 times SD (<DMSO> + 2 SD). (b) Sequences of long versions of active mimic peptides identified in initial screen.

Supplementary Figure. 3: Expression and purification of PFSGDS protein. SDS-PAGE gel of purified PFSGDS protein suggests the protein is >95% pure.

Supplementary Figure. 4: Minor structural differences in the JR5.1 TCR – HLA-DQ2.5 interface. Comparison of the of the interactions between the JR5.1 TCR and the HLA-DQ2.5 in the ternary complex with HLA-DQ2.5-*P.aerug-α2a* (left) and HLA-DQ2.5-glia-α2 (right). Colours: CDR loops and corresponding TCR footprint contacts are coloured red, pink, cyan, orange, purple and blue for CDR1α, CDR2α, CDR3α, CDR1β, CDR2β, CDR3β, respectively, and framework residues are coloured green. The peptide and HLA α- and β-chains are coloured grey, light green, and light yellow, respectively. (a) Comparison of interactions between the CDR3 region of the JR5.1 TCR and the HLA. (b) Comparison of germline encoded interactions between the JR5.1 TCR

and the HLA α-chain. (c) Comparison of germline encoded interactions between the JR5.1 TCR and the HLA β-chain.

Supplementary Figure. 5: Gating strategy for T cell stimulation assays. TCR transduced SKW3 cells were gated as follows: Lymphocytes (FCS-A/SSC-H); single cells (FSC-A/FSC-H); live cells (AQUA stain, BV525-A low); GFP expression (B530-A high). T cell activation was measured via staining of CD69 and CD3 (PE-anti-CD69, YG585-A vs. APC-anti-CD3, R670-A) and expressed as % of cells in the respective activation gate (CD69high and CD3low). Gates for T cell activation were set for each TCR transduced SKW3 line to include 5-10 % of cells in the vehicle control sample.

762 **Supplementary Table 1 – CeD cohort details for gluten challenges**

							Challenge symptoms (+ mild; ++ moderate; +++ severe)[c]									
Subject	**Age**	**Sex**	**HLA-DQ[a]**	**tTG-IgA**	**DGP-IgG**	**3d[b]**	**P**	**Bl**	**C**	**D**	**F**	**S**	**N**	**V**	**H**	**L**	**Other**
0670	61	F	DQ2.5/DQ8	17 (<20)[d]	44 (< 20)	Y		+					+				
0041	68	F	DQ2.5/DQX	ND	ND	Y	+										
0148	63	M	DQ2.5/DQX	3 (<20)	<3 (<20)	Y	+++	+++		+++	+++	+++			+	+++	
0570	61	M	DQ2.5/DQX	4 (<20)	<3 (<20)	Y											Asymptomatic
0496	43	F	DQ2.5/DQX	ND	ND	Y	+	+			+		++		++	+++	

[a] X denotes allele other than DQ2.5

[b] 3d indicates subjects completing all 3 days of gluten challenge (Y = YES, N = NO)

[c] N = nausea, Bl = bloating, V = vomiting, D = diarrhea, L = lethargy, P = pain/cramping, C = constipation, S = loose stool, F = flatulence, H = headaches

[d] numbers in brackets indicate the serological assay detection cut-off. Red result indicates a positive result.

ND = not done

763

764 References

765

766

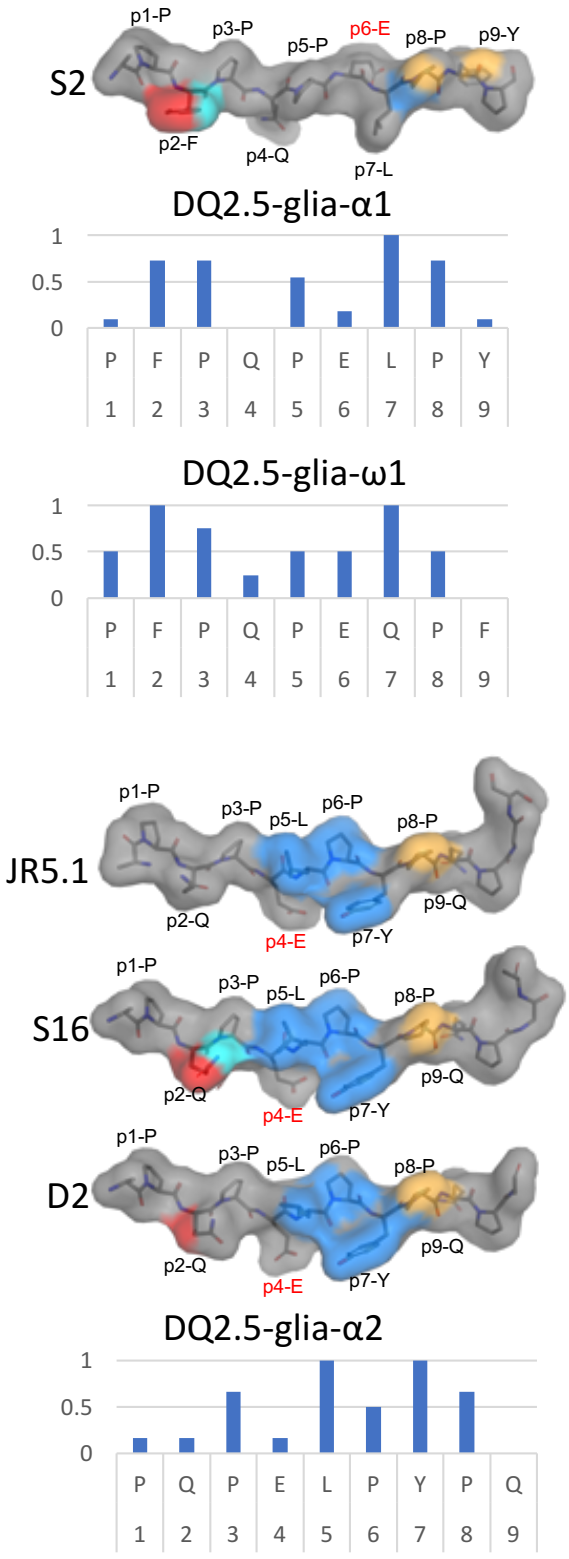
767 1. Trowsdale, J. & Knight, J.C. Major histocompatibility complex genomics and human disease.
768 *Annu Rev Genomics Hum Genet* **14**, 301-23 (2013).
769 2. Dendrou, C.A., Petersen, J., Rossjohn, J. & Fugger, L. HLA variation and disease. *Nat Rev*
770 *Immunol* **18**, 325-339 (2018).
771 3. Pollard, K.M., Hultman, P. & Kono, D.H. Toxicology of autoimmune diseases. *Chem Res*
772 *Toxicol* **23**, 455-66 (2010).
773 4. Marino, E. et al. Gut microbial metabolites limit the frequency of autoimmune T cells and
774 protect against type 1 diabetes. *Nat Immunol* **18**, 552-562 (2017).
775 5. Li, B., Selmi, C., Tang, R., Gershwin, M.E. & Ma, X. The microbiome and autoimmunity: a
776 paradigm from the gut-liver axis. *Cell Mol Immunol* **15**, 595-609 (2018).
777 6. Sollid, L.M. & Jabri, B. Triggers and drivers of autoimmunity: lessons from coeliac disease.
778 *Nat Rev Immunol* **13**, 294-302 (2013).
779 7. Sollid, L.M., Qiao, S.W., Anderson, R.P., Gianfrani, C. & Koning, F. Nomenclature and listing
780 of celiac disease relevant gluten T-cell epitopes restricted by HLA-DQ molecules.
781 *Immunogenetics* **64**, 455-60 (2012).
782 8. Tye-Din, J.A. et al. Comprehensive, quantitative mapping of T cell epitopes in gluten in
783 celiac disease. *Sci Transl Med* **2**, 41ra51 (2010).
784 9. Kim, C.Y., Quarsten, H., Bergseng, E., Khosla, C. & Sollid, L.M. Structural basis for HLA-DQ2-
785 mediated presentation of gluten epitopes in celiac disease. *Proc Natl Acad Sci U S A* **101**,
786 4175-9 (2004).
787 10. Henderson, K.N. et al. A structural and immunological basis for the role of human
788 leukocyte antigen DQ8 in celiac disease. *Immunity* **27**, 23-34 (2007).
789 11. Molberg, O. et al. Tissue transglutaminase selectively modifies gliadin peptides that are
790 recognized by gut-derived T cells in celiac disease. *Nat Med* **4**, 713-7 (1998).
791 12. Qiao, S.W. et al. Posttranslational modification of gluten shapes TCR usage in celiac
792 disease. *J Immunol* **187**, 3064-71 (2011).
793 13. Broughton, S.E. et al. Biased T cell receptor usage directed against human leukocyte
794 antigen DQ8-restricted gliadin peptides is associated with celiac disease. *Immunity* **37**, 611-
795 21 (2012).
796 14. Petersen, J. et al. T-cell receptor recognition of HLA-DQ2-gliadin complexes associated with
797 celiac disease. *Nat Struct Mol Biol* **21**, 480-8 (2014).
798 15. Qiao, S.W., Christophersen, A., Lundin, K.E. & Sollid, L.M. Biased usage and preferred
799 pairing of alpha- and beta-chains of TCRs specific for an immunodominant gluten epitope
800 in coeliac disease. *Int Immunol* **26**, 13-9 (2014).
801 16. Vriezinga, S.L. et al. Randomized Feeding Intervention in Infants at High Risk for Celiac
802 Disease. *New England Journal of Medicine* **371**, 1304-1315 (2014).
803 17. Marild, K., Kahrs, C.R., Tapia, G., Stene, L.C. & Stordal, K. Infections and risk of celiac
804 disease in childhood: a prospective nationwide cohort study. *Am J Gastroenterol* **110**,
805 1475-84 (2015).
806 18. Viitasalo, L. et al. Early microbial markers of celiac disease. *J Clin Gastroenterol* **48**, 620-4
807 (2014).
808 19. Ludvigsson, J.F. & Murray, J.A. Epidemiology of Celiac Disease. *Gastroenterol Clin North Am*
809 **48**, 1-18 (2019).

20. Kagnoff, M.F. et al. Evidence for the role of a human intestinal adenovirus in the pathogenesis of coeliac disease. *Gut* **28**, 995-1001 (1987).
21. Kagnoff, M.F., Austin, R.K., Hubert, J.J., Bernardin, J.E. & Kasarda, D.D. Possible role for a human adenovirus in the pathogenesis of celiac disease. *J Exp Med* **160**, 1544-57 (1984).
22. Dahal-Koirala, S. et al. Discriminative T-cell receptor recognition of highly homologous HLA-DQ2-bound gluten epitopes. *J Biol Chem* **294**, 941-952 (2019).
23. Purcell, A.W., Ramarathinam, S.H. & Ternette, N. Mass spectrometry-based identification of MHC-bound peptides for immunopeptidomics. *Nat Protoc* **14**, 1687-1707 (2019).
24. Anderson, R.P., Degano, P., Godkin, A.J., Jewell, D.P. & Hill, A.V.S. In vivo antigen challenge in celiac disease identifies a single transglutaminase-modified peptide as the dominant A-gliadin T-cell epitope. *Nat Med* **6**, 337-342 (2000).
25. Verdu, E.F., Galipeau, H.J. & Jabri, B. Novel players in coeliac disease pathogenesis: role of the gut microbiota. *Nat Rev Gastroenterol Hepatol* **12**, 497-506 (2015).
26. Stene, L.C. et al. Rotavirus infection frequency and risk of celiac disease autoimmunity in early childhood: a longitudinal study. *Am J Gastroenterol* **101**, 2333-40 (2006).
27. Bouziat, R. et al. Reovirus infection triggers inflammatory responses to dietary antigens and development of celiac disease. *Science* **356**, 44-50 (2017).
28. Kahrs, C.R. et al. Enterovirus as trigger of coeliac disease: nested case-control study within prospective birth cohort. *BMJ* **364**, l231 (2019).
29. Ashorn, S. et al. Elevated serum anti-Saccharomyces cerevisiae, anti-I2 and anti-OmpW antibody levels in patients with suspicion of celiac disease. *J Clin Immunol* **28**, 486-94 (2008).
30. Riddle, M.S., Murray, J.A. & Porter, C.K. The incidence and risk of celiac disease in a healthy US adult population. *Am J Gastroenterol* **107**, 1248-55 (2012).
31. Riddle, M.S., Murray, J.A., Cash, B.D., Pimentel, M. & Porter, C.K. Pathogen-specific risk of celiac disease following bacterial causes of foodborne illness: a retrospective cohort study. *Dig Dis Sci* **58**, 3242-5 (2013).
32. Sanchez, E., Donat, E., Ribes-Koninckx, C., Fernandez-Murga, M.L. & Sanz, Y. Duodenal-mucosal bacteria associated with celiac disease in children. *Appl Environ Microbiol* **79**, 5472-9 (2013).
33. Wacklin, P. et al. Altered duodenal microbiota composition in celiac disease patients suffering from persistent symptoms on a long-term gluten-free diet. *Am J Gastroenterol* **109**, 1933-41 (2014).
34. Viitasalo, L. et al. Microbial Biomarkers in Patients with Nonresponsive Celiac Disease. *Dig Dis Sci* **63**, 3434-3441 (2018).
35. D'Argenio, V. et al. Metagenomics Reveals Dysbiosis and a Potentially Pathogenic N. flavescens Strain in Duodenum of Adult Celiac Patients. *Am J Gastroenterol* **111**, 879-90 (2016).
36. Caminero, A. et al. Duodenal Bacteria From Patients With Celiac Disease and Healthy Subjects Distinctly Affect Gluten Breakdown and Immunogenicity. *Gastroenterology* **151**, 670-83 (2016).
37. Caminero, A. et al. Duodenal bacterial proteolytic activity determines sensitivity to dietary antigen through protease-activated receptor-2. *Nat Commun* **10**, 1198 (2019).
38. Dickson, R.P. et al. Cell-associated bacteria in the human lung microbiome. *Microbiome* **2**, 28 (2014).
39. N'Diaye, A. et al. Substance P and Calcitonin Gene-Related Peptide: Key Regulators of Cutaneous Microbiota Homeostasis. *Front Endocrinol (Lausanne)* **8**, 15 (2017).

40. Scales, B.S., Dickson, R.P., LiPuma, J.J. & Huffnagle, G.B. Microbiology, genomics, and clinical significance of the Pseudomonas fluorescens species complex, an unappreciated colonizer of humans. *Clin Microbiol Rev* **27**, 927-48 (2014).
41. Dalwadi, H., Wei, B., Kronenberg, M., Sutton, C.L. & Braun, J. The Crohn's disease-associated bacterial protein I2 is a novel enteric t cell superantigen. *Immunity* **15**, 149-58 (2001).
42. Sutton, C.L. et al. Identification of a novel bacterial sequence associated with Crohn's disease. *Gastroenterology* **119**, 23-31 (2000).
43. Segal, Y. & Shoenfeld, Y. Vaccine-induced autoimmunity: the role of molecular mimicry and immune crossreaction. *Cell Mol Immunol* **15**, 586-594 (2018).
44. Lang, H.L. et al. A functional and structural basis for TCR cross-reactivity in multiple sclerosis. *Nat Immunol* **3**, 940-3 (2002).
45. Cole, D.K. et al. Hotspot autoimmune T cell receptor binding underlies pathogen and insulin peptide cross-reactivity. *The Journal of Clinical Investigation* **126**, 3626-3626 (2016).
46. Kabsch, W. Xds. *Acta Crystallogr D Biol Crystallogr* **66**, 125-32 (2010).
47. Winn, M.D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**, 235-42 (2011).
48. McCoy, A.J. et al. Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674 (2007).
49. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501 (2010).
50. Adams, P.D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-21 (2010).
51. Gras, S. et al. Allelic polymorphism in the T cell receptor and its impact on immune responses. *J Exp Med* **207**, 1555-67 (2010).

Figure 1

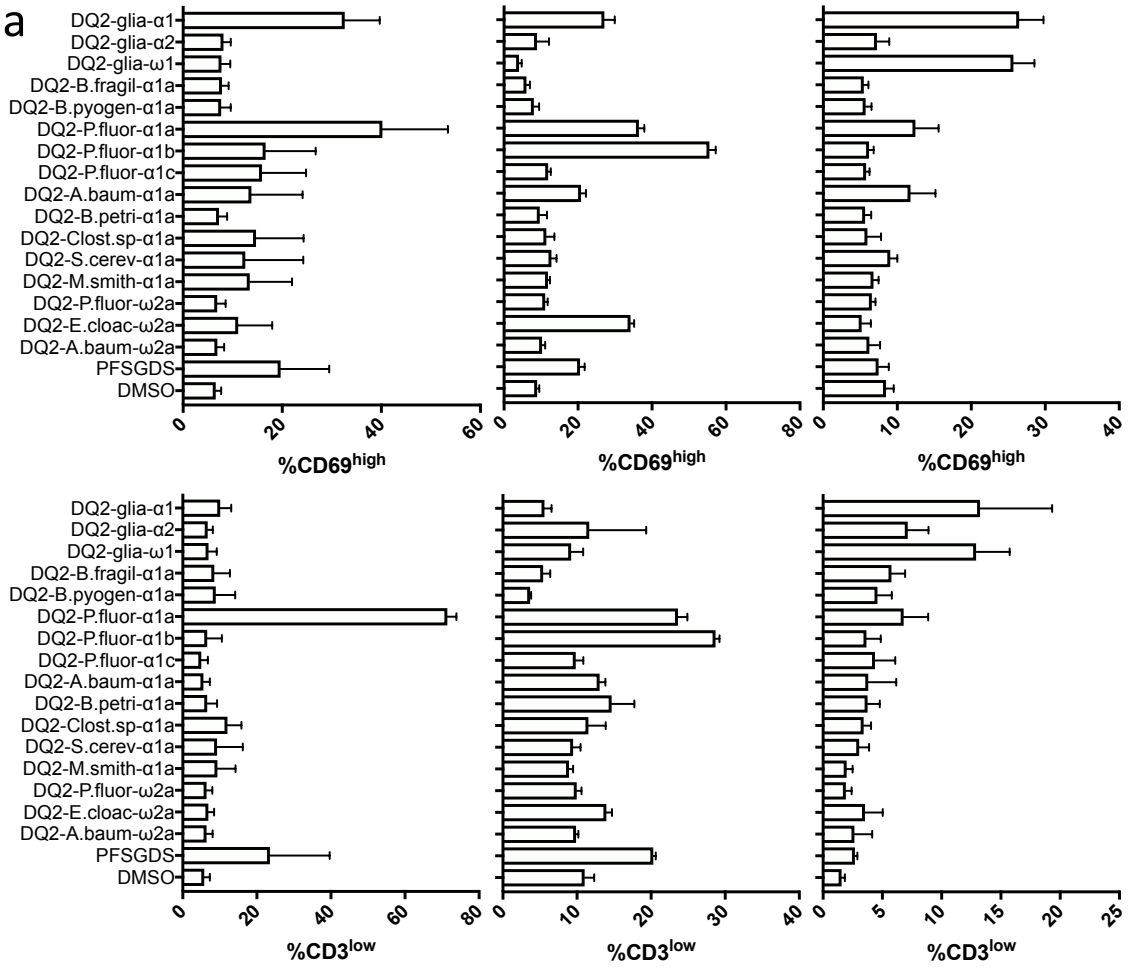
a



b

Peptide	sequence
DQ2.5-glia-α1	**LQPFPQPELPY**
DQ2.5-B.fragil-α1a	KELPQPELPYS
DQ2.5-B.pyogen-α1a	PPLPQPEVPET
DQ2.5-P.fluor-α1a	PMPMPELPYP
DQ2.5-P.fluor-α1b	PMPLPDLPYP
DQ2.5-P.fluor-α1c	VNYPHPDVPYT
DQ2.5-A.baum-α1.a	VQPWPQPELPEY
DQ2.5-B.petri-α1a	PYTLPELPYDA
DQ2.5-Clost.sp-α1a	LPYPQPDLPGV
DQ2.5-S.cerev-α1a	RYMPDPELPYI
DQ2.5-M.smith-α1a	DVFKVEPEIPY
DQ2.5-glia-ω1	**PFPQPEQPF**
DQ2.5-P.fluor-ω1a	PPLPEPEQPPV
DQ2.5-E.cloac-ω1a	PPFPEGEQPFP
DQ2.5-A.baum-ω1a	SQPIPQPEQPP
DQ2.5-glia-α2	**FPQPELPYPQP**
DQ2.5-P.fluor-α2a	PMPELPYPAT
DQ2.5-P.fluor-α2b	PLPELPYPAT
DQ2.5-P.fluor-α2c	NPPPDLPYPDI
DQ2.5-B.fragil-α2a	LPQPELPYSEM
DQ2.5-B.copro-α2a	LPLPDLPYPVA
DQ2.5-S.fonti-α2a	LPLPELPYSQP
DQ2.5-P.aerug-α2a	MVVQSELPYPE
DQ2.5-S.cerev-α2a	MPDPELPYINL
DQ2.5-M.smith-α2a	KVEPEIPYPED
DQ2.5-B.cereus-α2a	KPQPEQPKPQP

Figure 2:


a
LS2.8/3.15
N12
L6
DQ2-glia-α1
DQ2-glia-α2
DQ2-glia-ω1
DQ2-B.fragil-α1a
DQ2-B.pyogen-α1a
DQ2-P.fluor-α1a
DQ2-P.fluor-α1b
DQ2-P.fluor-α1c
DQ2-A.baum-α1a
DQ2-B.petri-α1a
DQ2-Clost.sp-α1a
DQ2-S.cerev-α1a
DQ2-M.smith-α1a
DQ2-P.fluor-ω2a
DQ2-E.cloac-ω2a
DQ2-A.baum-ω2a
PFSGDS
DMSO
%CD69high
%CD3low

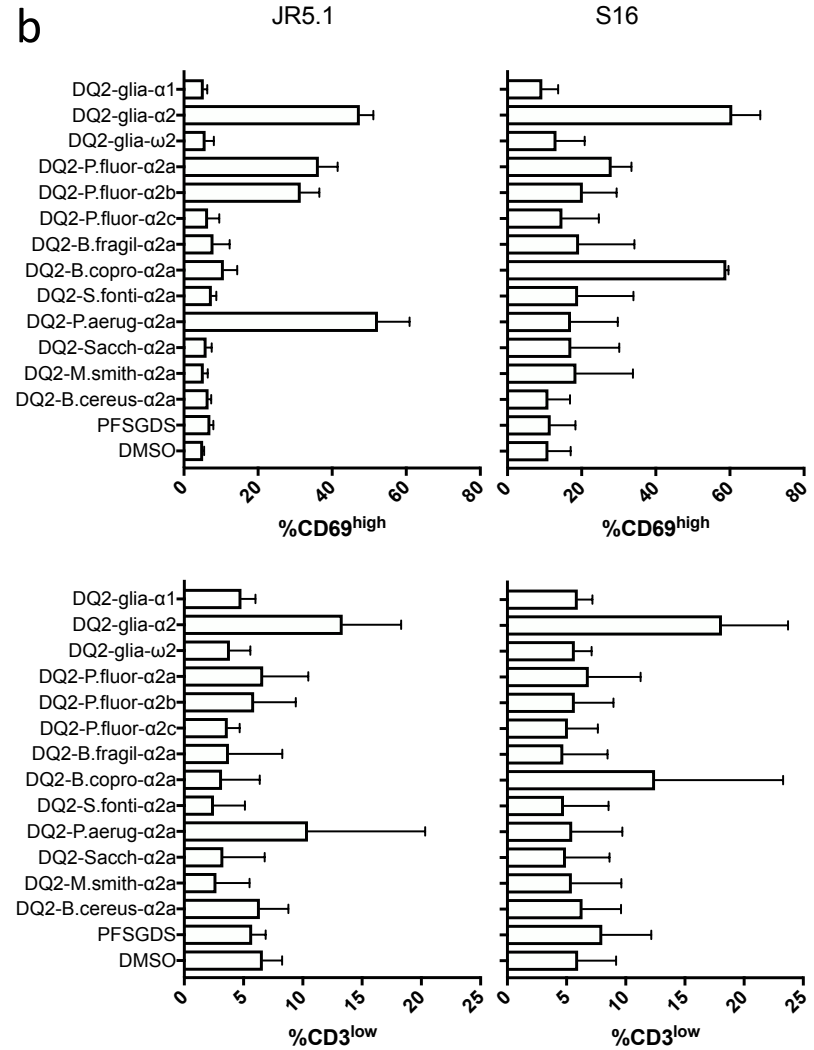
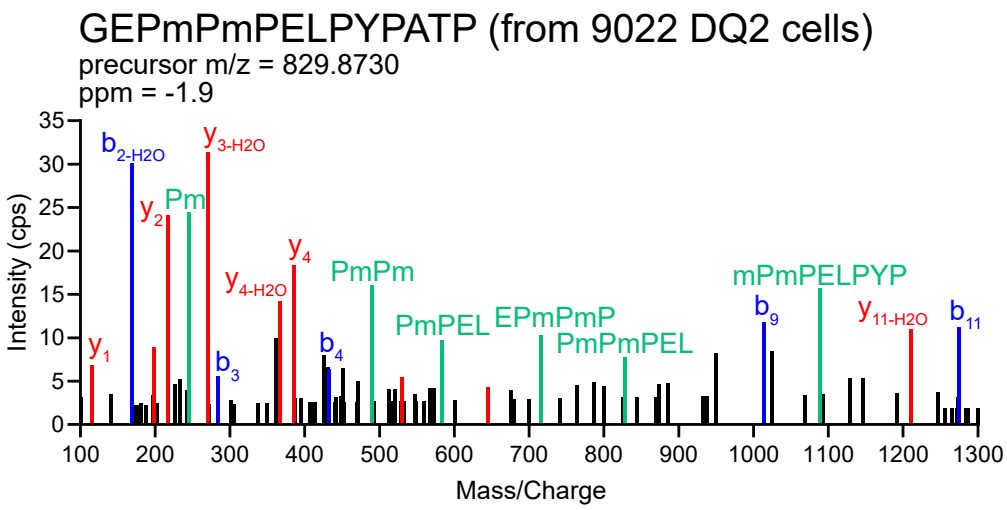

b
JR5.1
S16
DQ2-glia-α1
DQ2-glia-α2
DQ2-glia-ω2
DQ2-P.fluor-α2a
DQ2-P.fluor-α2b
DQ2-P.fluor-α2c
DQ2-B.fragil-α2a
DQ2-B.copro-α2a
DQ2-S.fonti-α2a
DQ2-P.aerug-α2a
DQ2-Sacch-α2a
DQ2-M.smith-α2a
DQ2-B.cereus-α2a
PFSGDS
DMSO
%CD69high
%CD3low

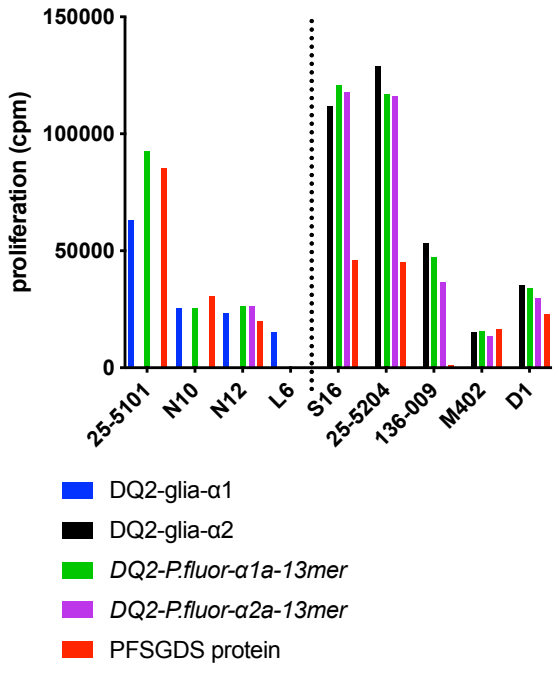
Figure3

a

b



c



d

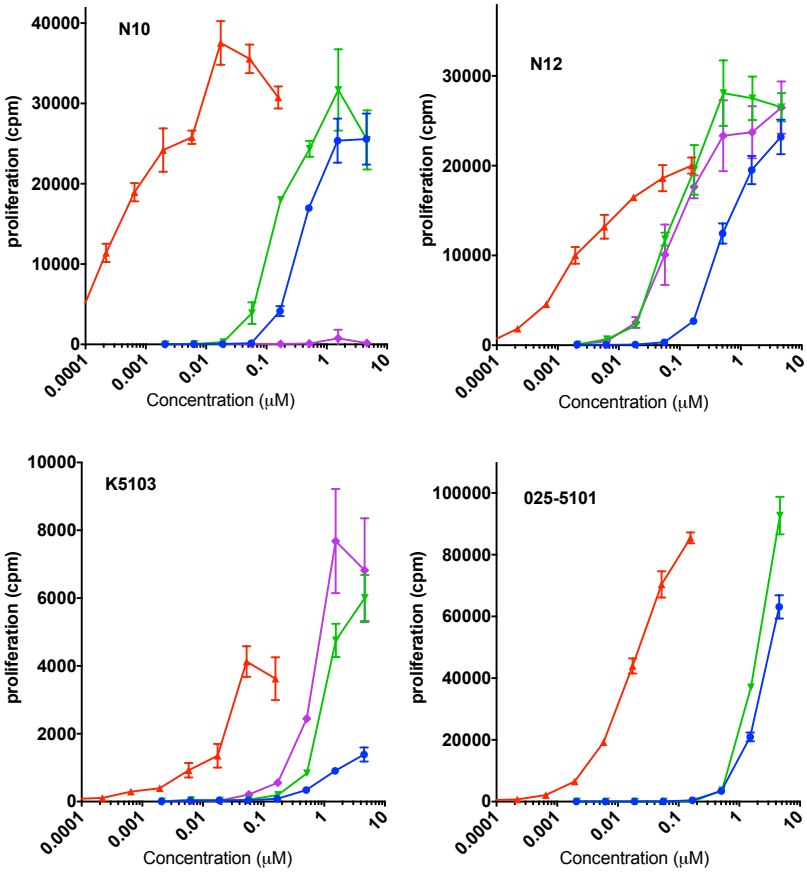


Figure 4

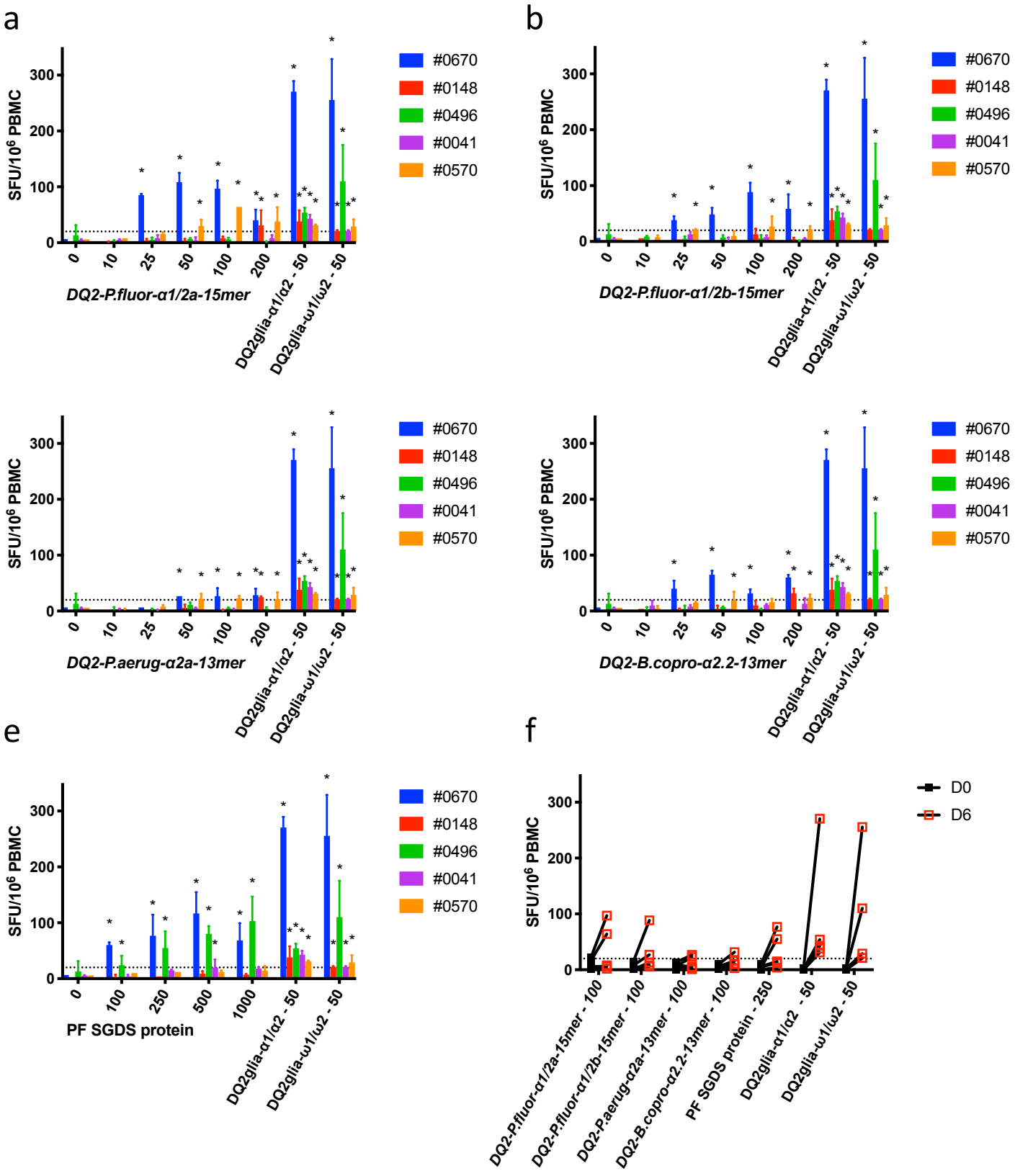

a
b
e
f
SFU/10^6 PBMC
#0670
#0148
#0496
#0041
#0570
DQ2-P.fluor-α1/2a-15mer
DQ2-P.fluor-α1/2b-15mer
DQ2-P.aerug-α2a-13mer
DQ2-B.copro-α2.2-13mer
PF SGDS protein
DQ2glia-α1/α2 - 50
DQ2glia-ω1/ω2 - 50
D0
D6

Figure 5

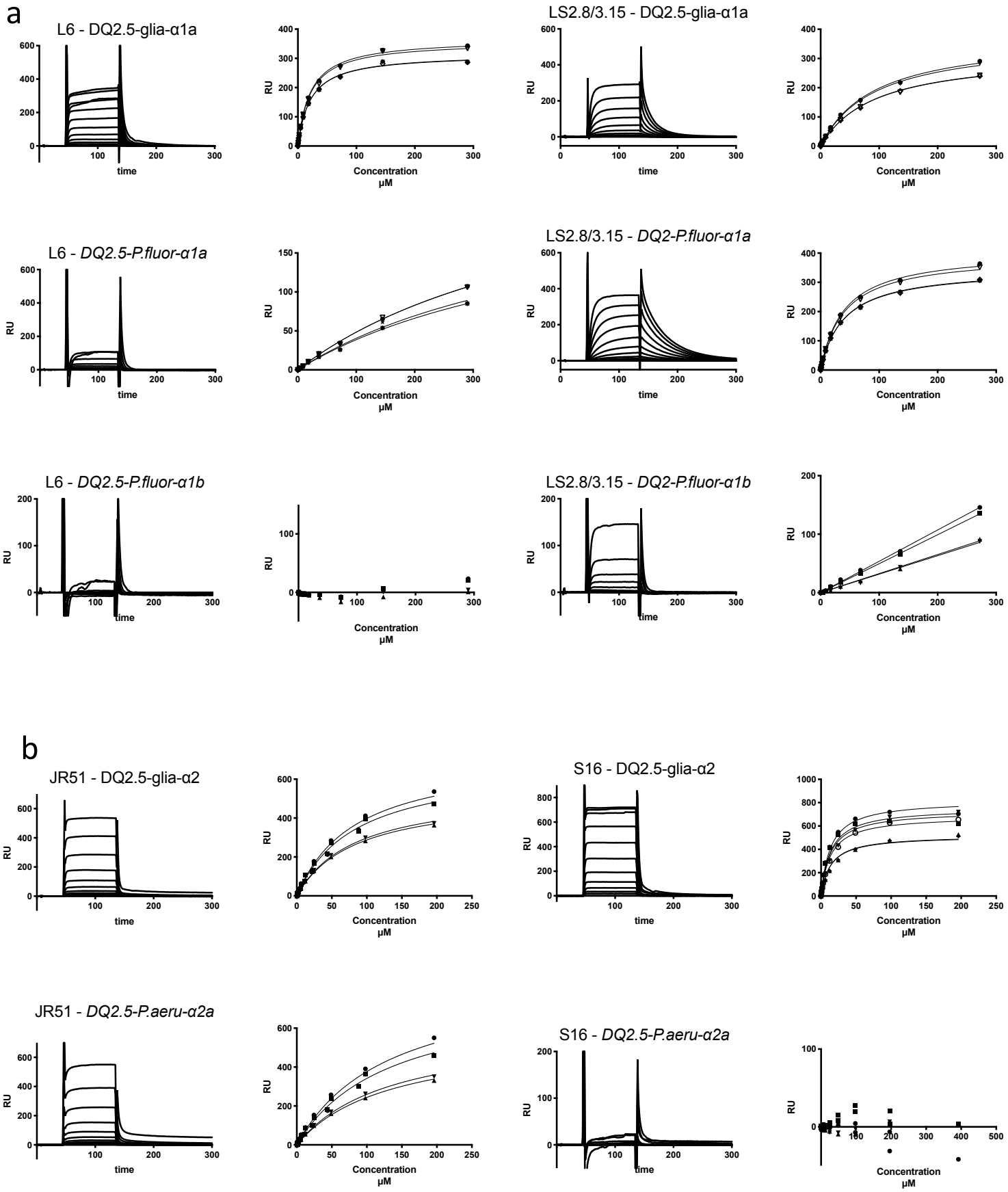
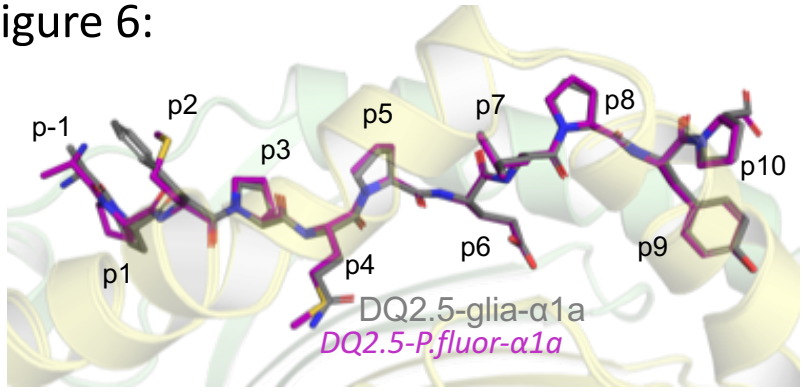
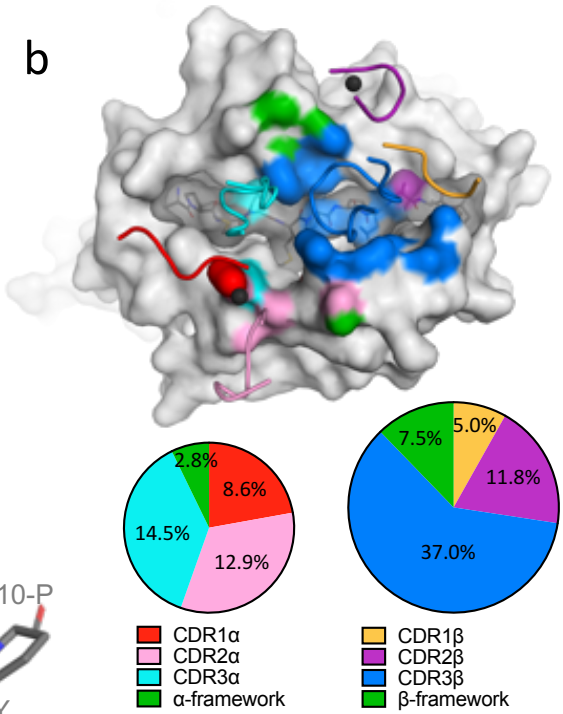
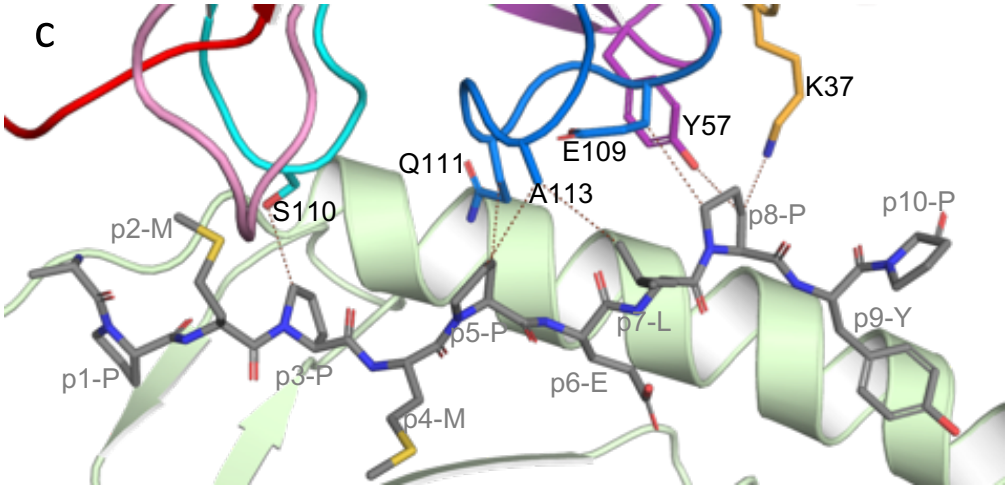
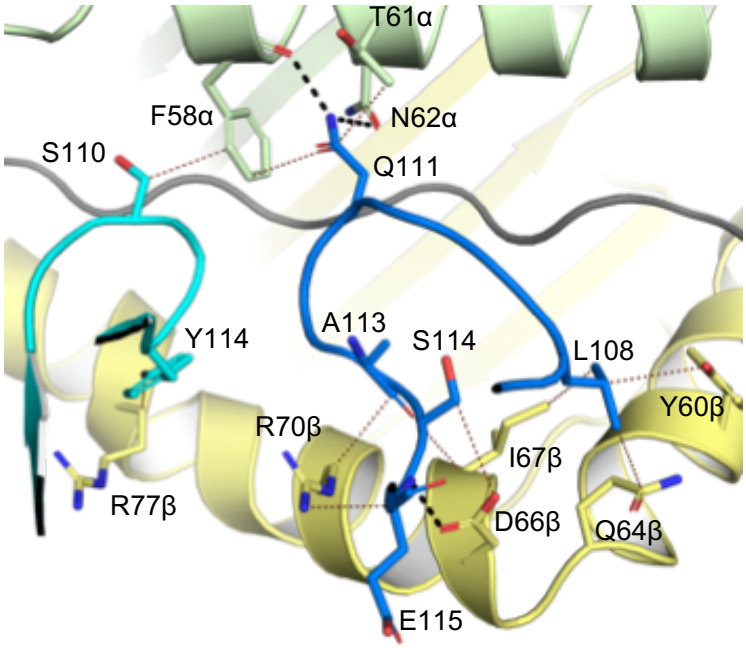

a
L6 - DQ2.5-glia-α1a
LS2.8/3.15 - DQ2.5-glia-α1a
L6 - DQ2.5-P.fluor-α1a
LS2.8/3.15 - DQ2-P.fluor-α1a
L6 - DQ2.5-P.fluor-α1b
LS2.8/3.15 - DQ2-P.fluor-α1b
b
JR51 - DQ2.5-glia-α2
S16 - DQ2.5-glia-α2
JR51 - DQ2.5-P.aeru-α2a
S16 - DQ2.5-P.aeru-α2a
RU
time
Concentration
μM

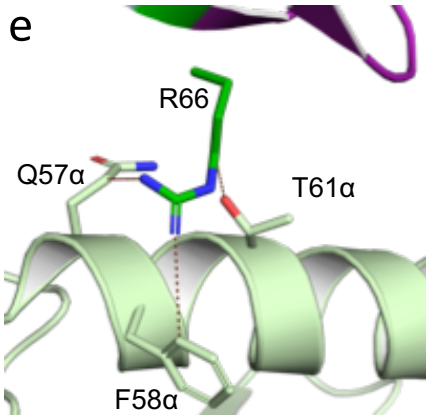
Figure 6:


a
p-1
p1
p2
p3
p4
p5
p6
p7
p8
p9
p10
DQ2.5-glia-α1a
DQ2.5-P.fluor-α1a


b
2.8%
8.6%
14.5%
12.9%
CDR1α
CDR2α
CDR3α
α-framework
7.5%
5.0%
11.8%
37.0%
CDR1β
CDR2β
CDR3β
β-framework


c
K37
Y57
E109
Q111
A113
S110
p2-M
p8-P
p10-P
p9-Y
p7-L
p5-P
p1-P
p3-P
p6-E
p4-M


d
T61α
F58α
N62α
S110
Q111
Y114
A113
S114
L108
Y60β
R70β
I67β
R77β
D66β
Q64β
E115


e
R66
Q57α
T61α
F58α

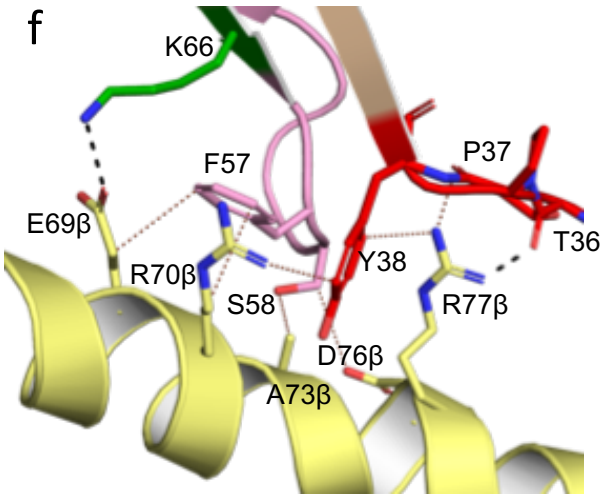
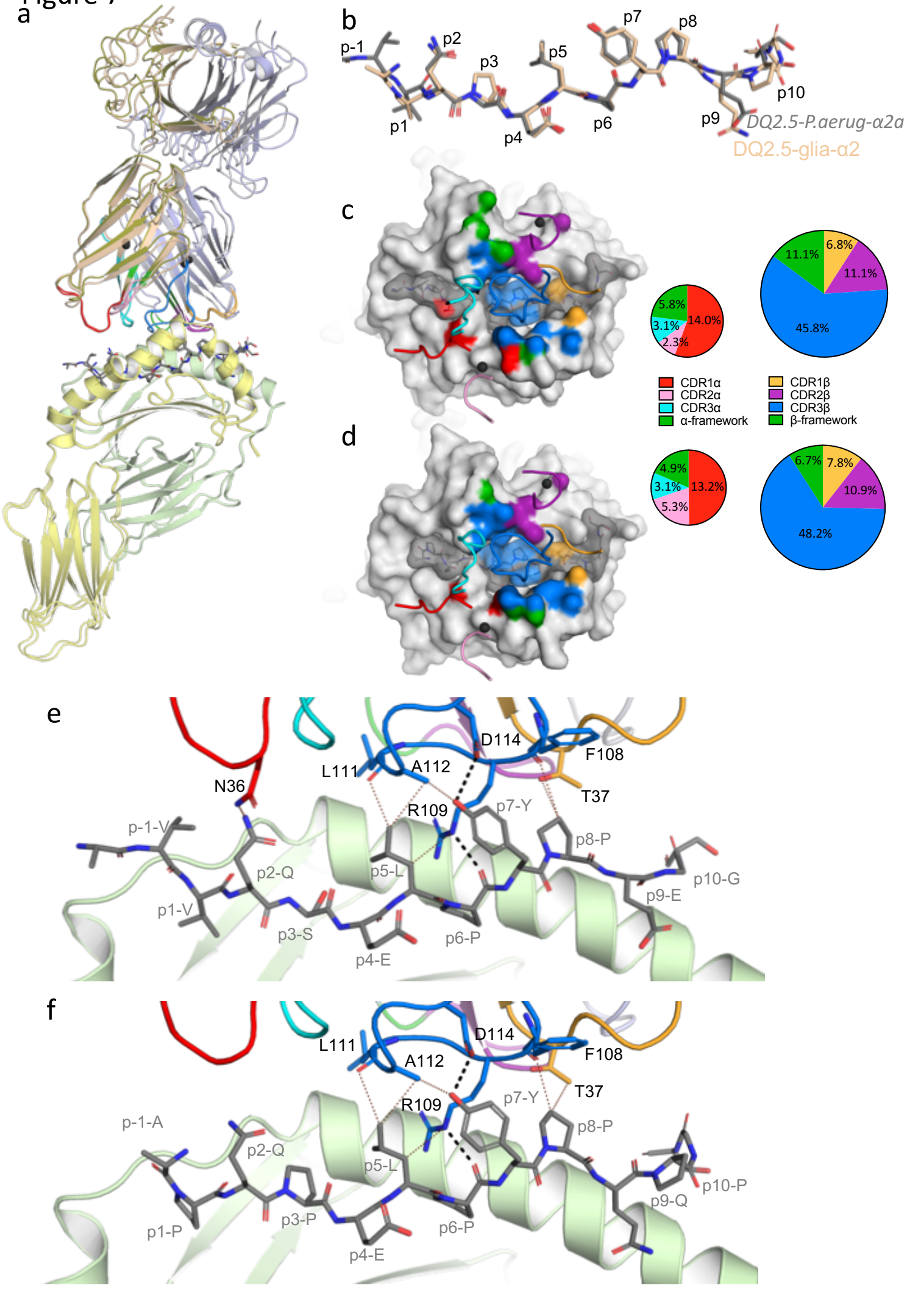
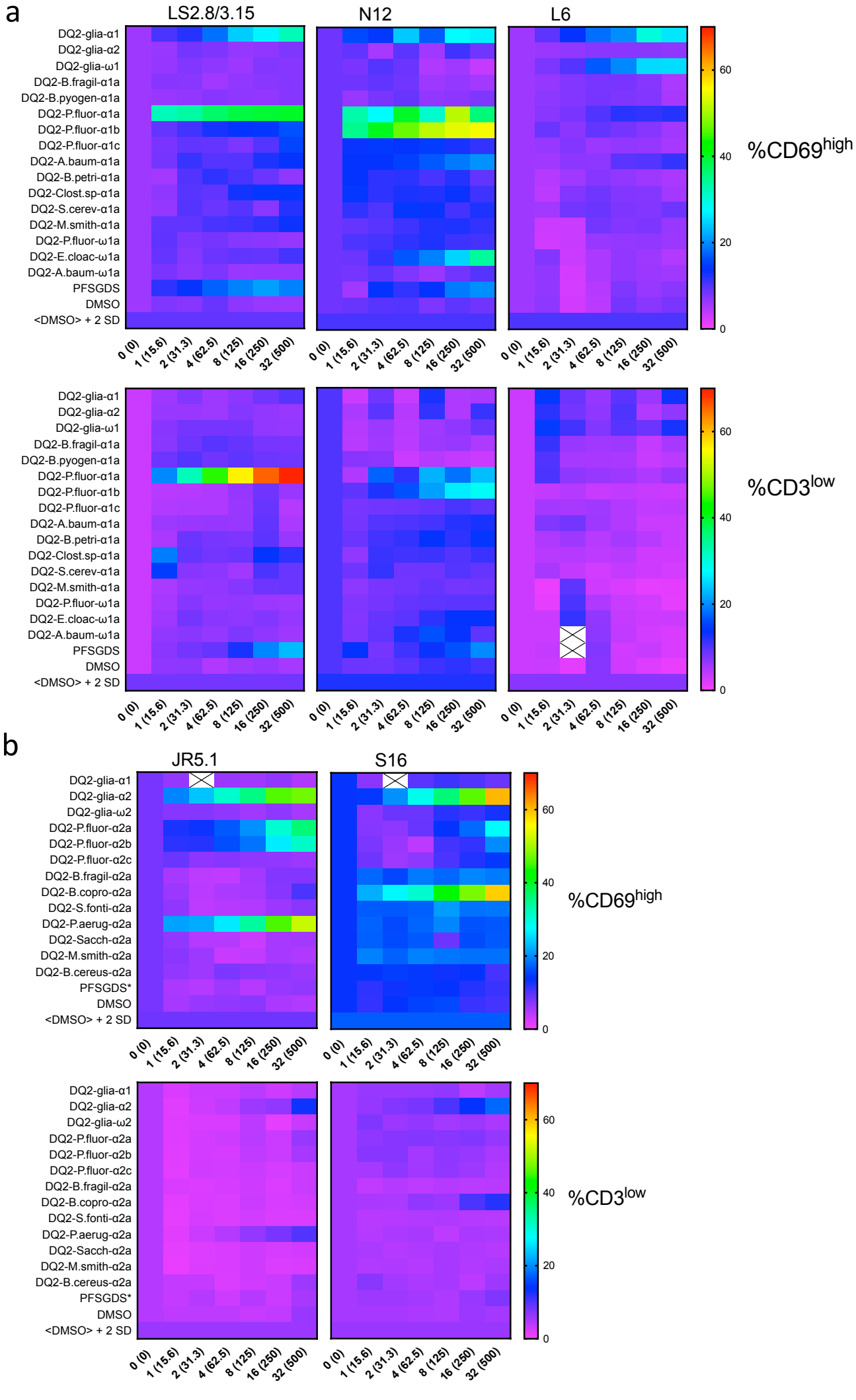

f
K66
F57
P37
E69β
T36
Y38
R70β
S58
R77β
D76β
A73β

Figure 7

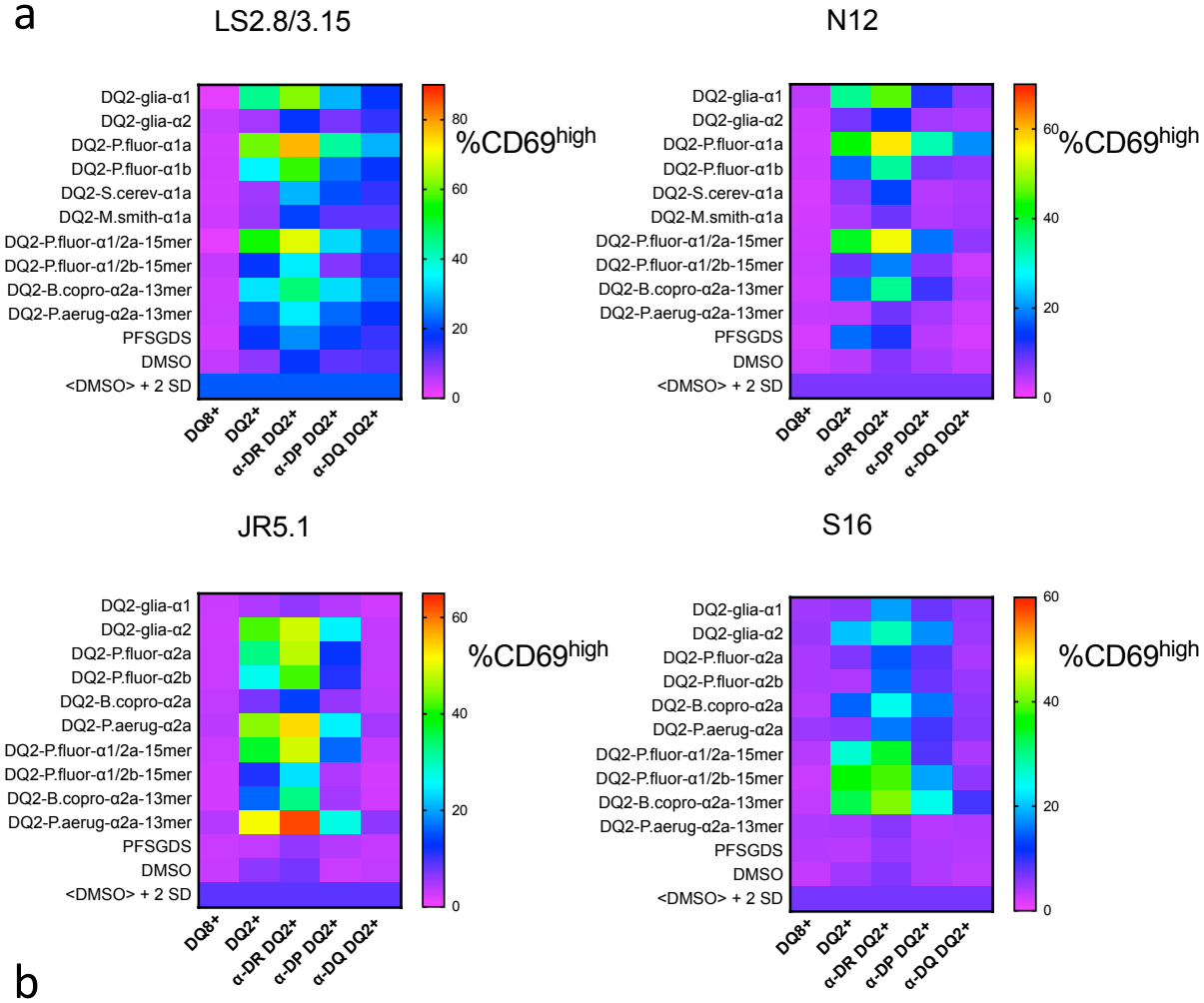

a
b
p-1
p1
p2
p3
p4
p5
p6
p7
p8
p9
p10
DQ2.5-P.aerug-α2a
DQ2.5-glia-α2
c
d
5.8%
14.0%
3.1%
2.3%
11.1%
6.8%
11.1%
45.8%
CDR1α
CDR2α
CDR3α
α-framework
CDR1β
CDR2β
CDR3β
β-framework
4.9%
13.2%
3.1%
5.3%
6.7%
7.8%
10.9%
48.2%
e
N36
L111
A112
D114
F108
T37
R109
p7-Y
p-1-V
p8-P
p2-Q
p5-L
p10-G
p9-E
p1-V
p3-S
p6-P
p4-E
f
L111
D114
A112
F108
T37
R109
p7-Y
p-1-A
p8-P
p2-Q
p5-L
p10-P
p9-Q
p1-P
p3-P
p6-P
p4-E

Supp. Figure 1



Supp. Figure 2

a

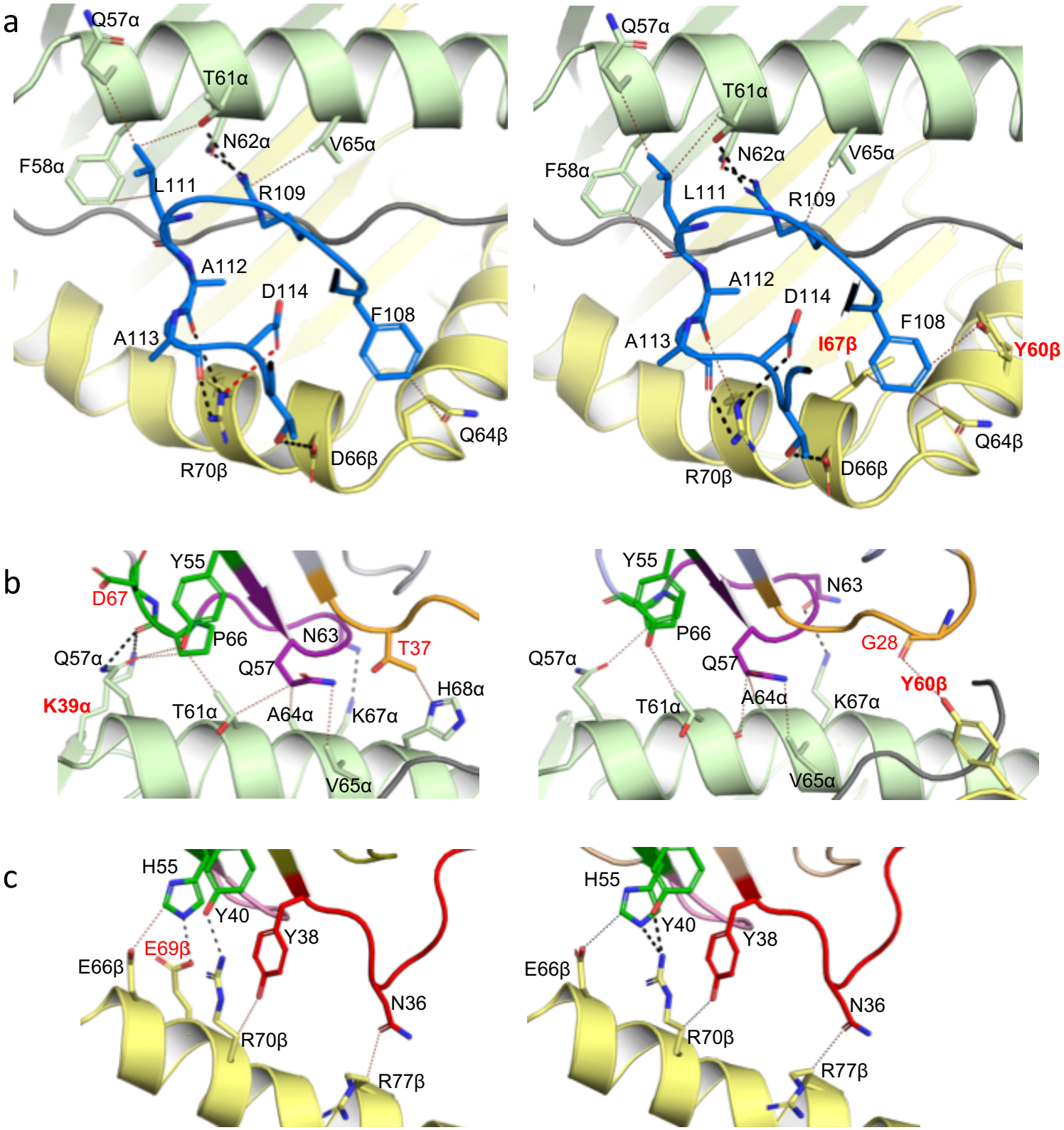


b

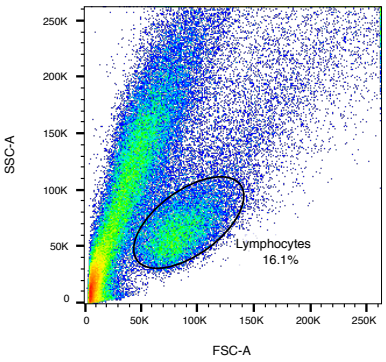
Peptide	sequence
DQ2.5-glia-α1	**LQPFPQPELPY**
DQ2.5-glia-α2	**FPQPELPYPQP**
DQ2.5-P.fluor-α1a	PMPMPELPYP
DQ2.5-P.fluor-α2a	PMPELPYPAT
DQ2.5-P.fluor-α1a-13mer	GEPMPMPELPYPA
DQ2.5-P.fluor-α2a-13mer	PMPMPELPYPATP
DQ2.5-P.fluor-α1/α2a-15mer	GEPMPMPELPYPATP
DQ2.5-P.fluor-α1b	PMPLPDLPYP
DQ2.5-P.fluor-α2c	PLPELPYPAT
DQ2.5-P.fluor-α1/α2b-15mer	GEPMPLPDLPYPATP
DQ2.5-P.aerug-α2a	MVVQSELPYPE
DQ2.5-P.aerug-α2a-13mer	MVVQSELPYPEGV
DQ2.5-B.copro-α2.2	LPLPDLPYPVA
DQ2.5-B.copro-α2.2-13mer	WLPLPDLPYPVAY
DQ2.5-glia-α1a/α2-15mer	**LQPFPQPELPYPQPQ**
DQ2.5-glia-ω1/ω2 -15mer	**QPFPQPEQPFPWQP**

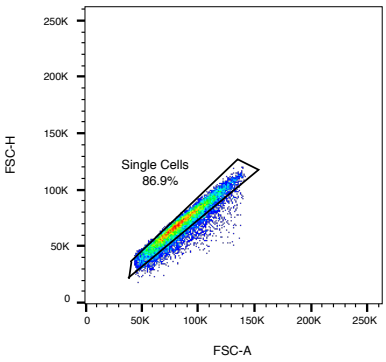
Supp. Figure 3

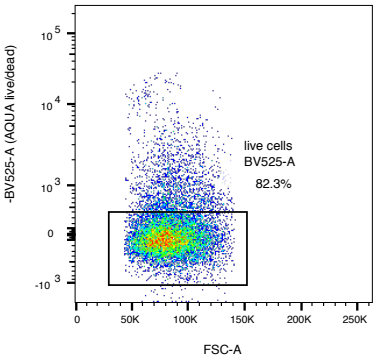
Supp. Figure 4

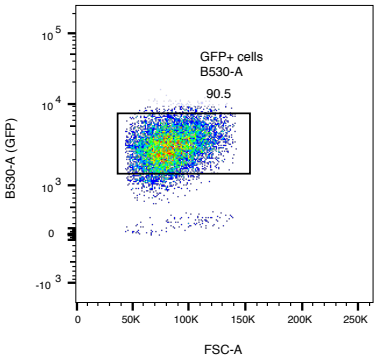


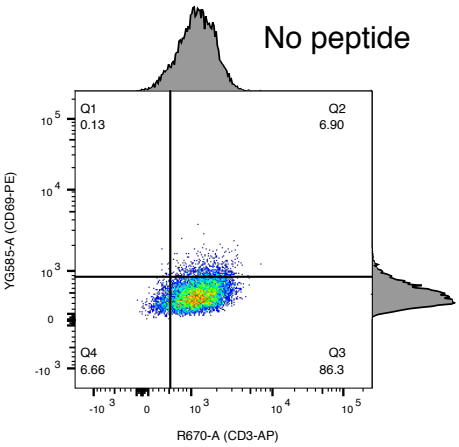
Supp. Figure 5



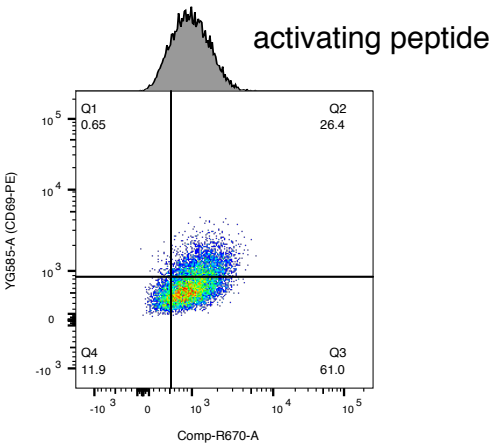








E11
FSC-A, Comp-B530-A subset
8257



C02
FSC-A, Comp-B530-A subset
8934