

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/119800/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Dahal-Koirala, Shiva, Ciacchi, Laura, Petersen, Jan, Risnes, Louise Fremgaard, Neumann, Ralf Stefan, Christophersen, Asbjørn, Lundin, Knut E. A., Reid, Hugh H., Qiao, Shuo-Wang, Rossjohn, Jamie and Sollid, Ludvig M. 2019. Discriminative T-cell receptor recognition of highly homologous HLA-DQ2-bound gluten epitopes. *Journal of Biological Chemistry* 294 (3) , pp. 941-952. 10.1074/jbc.RA118.005736

Publishers page: <http://dx.doi.org/10.1074/jbc.RA118.005736>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



## Discriminative T-cell receptor recognition of highly homologous HLA-DQ2-bound gluten epitopes

Shiva Dahal-Koirala<sup>1,2,§</sup>, Laura Ciacchi<sup>3,4,§</sup>, Jan Petersen<sup>3,4</sup>, Louise Fremgaard Risnes<sup>1,2</sup>, Ralf Stefan Neumann<sup>2</sup>, Asbjørn Christophersen<sup>2</sup>, Knut E.A. Lundin<sup>2,5</sup>, Hugh H. Reid<sup>3,4</sup>, Shuo-Wang Qiao<sup>1,2</sup>, Jamie Rossjohn<sup>3,4,6,\*</sup> & Ludvig M. Sollid<sup>1,2,\*</sup>

From the <sup>1</sup>Department of Immunology, University of Oslo and Oslo University Hospital-Rikshospitalet, 0372, Oslo, Norway; <sup>2</sup>K. G. Jebsen Centre for Coeliac Disease Research, University of Oslo, 0424, Oslo, Norway; <sup>3</sup>Infection and Immunity Program and Department of Biochemistry and Molecular Biology, Biomedicine Discovery Institute, Monash University, Clayton, Victoria 3800, Australia; <sup>4</sup>ARC Centre of Excellence in Advanced Molecular Imaging, Monash University, Clayton, Victoria 3800, Australia; <sup>5</sup>Department of Gastroenterology, Oslo University Hospital-Rikshospitalet, 0372, Oslo, Norway; <sup>6</sup>Institute of Infection and Immunity, School of Medicine, Cardiff University, Cardiff CF14 4XN, United Kingdom; <sup>§</sup>Shared first authors; <sup>\*</sup>Joint senior and corresponding authors.

Running title: *Discriminative TCR recognition of similar gluten epitopes*

To whom correspondence should be addressed:

Jamie Rossjohn : Infection and Immunity Program and Department of Biochemistry and Molecular Biology, Biomedicine Discovery Institute, Monash University, Clayton, Victoria 3800, Australia; [Jamie.rossjohn@monash.edu](mailto:Jamie.rossjohn@monash.edu)

Ludvig M. Sollid: Department of Immunology, University of Oslo and Oslo University Hospital-Rikshospitalet, 0372, Oslo, Norway; [l.m.sollid@medisin.uio.no](mailto:l.m.sollid@medisin.uio.no)

Keywords: human, T-cell, T-cell receptor (TCR), Celiac disease (CeD), gluten, gliadin, gluten intolerance, immune response, major histocompatibility complex, immunodominant epitope

### Abstract

Celiac disease (CeD) provides an opportunity to study the specificity underlying human T-cell responses to an array of similar epitopes presented by the same human leukocyte antigen-II (HLA-II) molecule. Here, we investigated T-cell responses to the two immunodominant and highly homologous HLA-DQ2.5-restricted gluten epitopes, DQ2.5-glia- $\alpha$ 1a (PFQPPELPY) and DQ2.5-glia- $\omega$ 1 (PFQPPEQPF). Using HLA-DQ2.5:DQ2.5-glia- $\alpha$ 1a and HLA-DQ2.5:DQ2.5-glia- $\omega$ 1 tetramers and single-cell  $\alpha\beta$  T-cell receptor (TCR) sequencing, we observed that despite similarity in biased variable-gene usage in the TCR repertoire responding to these nearly identical peptide:HLA-II complexes, most of the T cells are specific for either of the two epitopes. To understand the molecular basis of this exquisite fine specificity, we undertook Ala substitution assays revealing that the p7

residue (Leu/Gln) is critical for specific epitope recognition by both DQ2.5-glia- $\alpha$ 1a and DQ2.5-glia- $\omega$ 1-reactive T-cell clones. We determined high-resolution binary crystal structures of HLA-DQ2.5 bound to DQ2.5-glia- $\alpha$ 1a (2.0 Å) and DQ2.5-glia- $\omega$ 1 (2.6 Å). These structures disclosed that differences around the p7 residue subtly alter the neighboring sub-structure and electrostatic properties of the HLA-DQ2.5-peptide complex, providing the fine specificity underlying the responses against these two highly homologous gluten epitopes. This study underscores the ability of TCRs to recognize subtle differences in the peptide:HLA-II landscape in a human disease setting.

### Introduction

The  $\alpha\beta$  T-cell antigen receptor (TCR) recognizes peptide when presented by a given Major Histocompatibility Complex (MHC) molecule

(1). In celiac disease (CeD), an autoimmune-like and HLA-associated disorder, disease-driving T cells recognize deamidated cereal gluten peptides when presented by HLA-DQ2.5, HLA-DQ8 or HLA-DQ2.2 (2). Over 90 % of the CeD patients express HLA-DQ2.5, while the remainder express HLA-DQ8 or HLA-DQ2.2. The gluten proteome is extremely complex, comprising hundreds of proteins (termed  $\alpha$ -,  $\gamma$ -, and  $\omega$ -gliadins and high- and low-molecular weight glutenins) (3). This system is remarkable in that there is natural immune exposure to a vast array of similar peptide sequences, and strikingly, the subjects who develop CeD make T-cell responses to a limited number of epitopes, many of which share substantial homology (4). In HLA-DQ2.5 associated CeD, most patients respond to DQ2.5-glia- $\alpha$ 1a, DQ2.5-glia- $\alpha$ 2, DQ2.5-glia- $\omega$ 1 and DQ2.5-glia- $\omega$ 2 epitopes (5), while most patients expressing HLA-DQ8 respond to DQ8-glia- $\alpha$ 1 (6). Hence, these epitopes are termed immunodominant epitopes. T cells reactive with immunodominant epitopes are persistent for decades in the patients (7). Such T cells, expressing gut homing markers, can be isolated from gut biopsies and from peripheral blood albeit at much lower frequency in blood (7).

Studies of TCRs recognizing immunodominant HLA-DQ2.5 and HLA-DQ8 restricted gluten epitopes have demonstrated biased usage of variable (V)-gene segments (6,8-13). DQ2.5-glia- $\alpha$ 2-specific TCRs exhibit repeated gene usage of *TRAV26-1/TRBV7-2* with a conserved CDR3 $\beta$  motif harboring a non-germline-encoded Arg residue (8,9,12). The TCRs specific for DQ2.5-glia- $\omega$ 2 also display biased usage of *TRAV4* and *TRBV4* gene segments (12). DQ2.5-glia- $\alpha$ 1a-specific TCRs demonstrate biased expression of *TRAV4* and *TRBV20-1* or *TRBV29-1* genes (10). Similarly, the majority of DQ8-glia- $\alpha$ 1-specific T cells express *TRBV9* preferentially paired with *TRBV26-2*, expressing a non-germline encoded Arg residue in the CDR3 $\beta$  loop (6,11,13). The *TRBV9*-negative T cells specific to DQ8-glia- $\alpha$ 1 also show biased usage of *TRAV8-3/TRBV6-1*<sup>+</sup> TCRs, also with a conserved non-germline encoded Arg residue in the CDR3 $\beta$  loop (11).

Recently, structural studies have provided molecular insight into TCR recognition of gluten epitopes when presented by HLA-DQ8 or HLA-DQ2.5 (6,10,11,13). Three crystal structures of

*TRAV26-1/TRBV7-2*-TCR:HLA-DQ2.5:DQ2.5-glia- $\alpha$ 2 complexes revealed that the *TRBV7-2* bias is mostly attributed to the interactions mediated by the CDR3 $\beta$  loop and that the conserved non-germline encoded Arg in this loop serves as a lynchpin in the peptide:HLA-II interaction (10). A crystal structure of *TRAV4/TRBV20-1* TCR:HLA-DQ2.5:DQ2.5-glia- $\alpha$ 1a complex uncovered that the *TRAV4* bias is a consequence of interactions between the germline-encoded TCR residues and HLA-DQ2.5 (10). Crystal structures of *TRAV26-2/TRBV9*-TCR:HLA-DQ8:DQ8- $\alpha$ 1-gliadin and *TRAV20/TRBV9*-TCR:HLA-DQ8:DQ8- $\alpha$ 1-gliadin ternary complexes revealed that germline-encoded residues interacting with the HLA-DQ8 molecule and gliadin peptide residues formed the basis for the *TRBV9* bias (6,11,13). These two crystal structures together with a crystal structures of a complex of *TRAV8-3/TRBV6-1*-TCR:HLA-DQ8:DQ8- $\alpha$ 1-gliadin (11) revealed that diverse TCR gene usage by DQ8-restricted gluten-specific T cells converges into a consensus binding that is mediated by germline or non-germline encoded Arg residue. However, it remains unclear whether highly homologous gluten epitopes would generate cross-reactive or highly specific T cell responses.

Here, we have studied the T-cell response towards the epitopes DQ2.5-glia- $\alpha$ 1a (PFQPELPY) and DQ2.5-glia- $\omega$ 1 (PFQPEQPF). These immunodominant gluten epitopes in CeD are highly homologous, with only two amino acid differences in the 9-mer core region. Despite the epitope similarity, we found that the majority of patient-derived T cells, both isolated from blood or the celiac gut lesion, are specific for either of the two epitopes with limited cross-reactivity. This suggests that there are important differences in how T cells recognize these peptide:HLA-II complexes, as the immune system of the patients are exposed simultaneously to both epitopes when they consume gluten-containing food. To explore the basis for specific recognition of the homologous epitopes, we performed T-cell proliferation assays with amino-acid substituted epitopes, undertook single-cell TCR $\alpha\beta$ -gene sequencing and solved the crystal structures of HLA-DQ2.5 complexed with either DQ2.5-glia- $\alpha$ 1a or DQ2.5-glia- $\omega$ 1. Despite biased usage of common V genes, the majority of the T cells were specific to only one epitope. The p7 residue in both epitopes was critical for the specific TCR

recognition by the discriminatory T-cell clones (TCCs). The crystal structures of the two peptide:HLA-DQ2.5 complexes, whilst similar, exhibited local structural perturbations around the p7 residue. Hence, this study demonstrates the ability of TCR to distinguish subtle differences in peptide:HLA-II topology. Consequently, in human T-cell mediated diseases like CeD, discrete alterations in the peptide:HLA-II landscape can profoundly shape the disease-driven immune response.

## Results

### Majority of T cells in blood and gut of CeD patients specific for DQ2.5-glia- $\alpha$ 1a or DQ2.5-glia- $\omega$ 1 show exquisite fine specificity

We identified distinct populations of CD4<sup>+</sup> T cells that bound the HLA-DQ2.5:DQ2.5-glia- $\alpha$ 1a or HLA-DQ2.5:DQ2.5-glia- $\omega$ 1 tetramers in blood and gut of 12 CeD patients. (Fig. 1A). The relative ratio of CD4<sup>+</sup> T cells positive to both DQ2.5:DQ2.5-glia- $\alpha$ 1a- and DQ2.5:DQ2.5-glia- $\omega$ 1-tetramers with CD4<sup>+</sup> T cells that are positive to either of the two tetramers in samples obtained from 11 CeD patients was 0.06 (mean). This suggests that the majority (~ 95 %) of these CD4<sup>+</sup> T cells recognize either of the two tetramers (Fig. 1B, Table S1) indicating that despite only two amino acids difference in the 9-mer core region of the two gliadin peptides, most of the *in vivo* primed T cells in blood and gut of CeD patients discriminate between the two epitopes.

The HLA-DQ:gluten tetramers (including HLA-DQ2.5:DQ2.5-glia- $\alpha$ 1a and HLA-DQ2.5:DQ2.5-glia- $\omega$ 1) give staining of effector memory CD4 T cells of CeD patients, but not of healthy subjects (14,15). We validated the specific staining of HLA-DQ2.5:DQ2.5-glia- $\alpha$ 1a vs HLA-DQ2.5:DQ2.5-glia- $\omega$ 1 tetramers with TCCs that were generated from tetramer-sorted cells by antigen-free expansion and cloning by limited dilution. On re-testing of antigen specificity, all (nine) TCCs generated from PBMCs of the patient CD1383, stained with the tetramer used for sorting and showed a proliferative response towards the respective peptide (Fig. 1C and Fig. 1D). One of the five TCCs generated from HLA-DQ2.5:DQ2.5-glia- $\alpha$ 1a tetramer binding CD4<sup>+</sup> T cells and one of four of the TCCs generated from HLA-DQ2.5:DQ2.5-glia- $\omega$ 1 tetramer binding CD4<sup>+</sup> T cells stained with both the tetramers and gave a response to both

peptides in T-cell proliferation assays. The cross-reactive T-cell clones displayed a higher fluorescence intensity to the tetramer originally used for their isolation. None of the nine TCCs stained with the other HLA-DQ2.5:gluten tetramers (HLA-DQ2.5:DQ2.5-glia- $\alpha$ 2, HLA-DQ2.5:DQ2.5-glia- $\omega$ 2 and CLIP2), indicating the antigen specificity of the tetramer staining (Fig. 1C). In addition, none of these TCCs showed response to any other gluten epitopes when tested in a T-cell proliferation assay against a panel of known gluten peptides at a high concentration of 10  $\mu$ M (Fig. 1D). Taken together, the results of the HLA-DQ:gluten tetramer staining of T cells in blood and gut as well as the proliferation assay and HLA-DQ:gluten tetramer staining of established T-cell clones (the low number here prevents an exact estimate of cross-reactivity), strongly indicate that T cells of CeD patients generally discriminate between the two epitopes.

### The p7 residue of the DQ2.5-glia- $\alpha$ 1a and DQ2.5-glia- $\omega$ 1 epitopes is critical for T-cell recognition

To explore the differences in recognition of these two highly homologous epitopes we measured the effect of single Ala substitutions at each position in the two epitopes (Table 1) in a T-cell proliferation assay. The sequences of the peptides used in these T-cell proliferation assays were identical to the epitope sequences that were part of the HLA-DQ:gluten tetramers. In this analysis TCCs that stained with the HLA-DQ2.5:DQ2.5-glia- $\alpha$ 1a or HLA-DQ2.5:DQ2.5-glia- $\omega$ 1 tetramers and that had reactivity with only one of the two epitopes, were tested (Table 2). Reactivity pattern of the DQ2.5-glia- $\alpha$ 1a-specific- and DQ2.5-glia- $\omega$ 1-specific TCCs to the single Ala-substituted DQ2.5-glia- $\alpha$ 1a and DQ2.5-glia- $\omega$ 1 peptides were normalized to the response to the wild-type peptide.

Both DQ2.5-glia- $\alpha$ 1a-(Fig. 2A) and DQ2.5-glia- $\omega$ 1-specific (Fig. 2B) TCCs lost reactivity on Ala substitution at p2 and p7. However, substitution at p9, a position that differs between the two epitopes, had no effect on the reactivity of the TCCs. This is in line with the observations from a previous study where p7 was critical, p9 was dispensable and the p2 had variable effects on T-cell recognition of the DQ2.5-glia- $\alpha$ 1a epitopes by the DQ2.5-glia- $\alpha$ 1a-specific TCCs (10). Unique to all DQ2.5-glia- $\alpha$ 1a-specific TCCs was the loss of reactivity on Ala substitution of

residues p6 and p8. Two of the three TCCs lost reactivity on substitution at p3 and p4 while the third TCC was sensitive to substitution at p5. DQ2.5-glia- $\omega$ 1-specific TCCs exhibited varying responses to Ala substitutions at positions other than p2 and p7. Substitution at p3 and p8 resulted in loss of reactivity in three of four TCCs. In brief, the p2 and p7 amino acids were important for T-cell recognition of both the DQ2.5-glia- $\alpha$ 1a and DQ2.5-glia- $\omega$ 1 epitopes.

To further investigate the basis of TCR specificity, T-cell proliferative assays were used to analyze the reactivity of DQ2.5-glia- $\alpha$ 1a- and DQ2.5-glia- $\omega$ 1-specific (Fig. 2C and Fig. 2D) TCCs against chimeric peptide with a single amino acid exchange at position p7 and p9 of DQ2.5-glia- $\alpha$ 1a ( $\alpha$ 1\_Qp7L,  $\alpha$ 1\_Fp9Y) and DQ2.5-glia- $\omega$ 1 ( $\omega$ 1\_Lp7Q,  $\omega$ 1\_Yp9F) (Table 3). Since the 9-mer core sequences of the DQ2.5-glia- $\alpha$ 1a and DQ2.5-glia- $\omega$ 1 differ only at p7 and p9, the pair of chimeras  $\alpha$ 1\_Qp7L and  $\omega$ 1\_Yp9F, as well as the pair of  $\alpha$ 1\_Fp9Y and  $\omega$ 1\_Lp7Q, will have identical 9-mer core sequence but different flanking sequences. We observed that the responses of DQ2.5-glia- $\alpha$ 1a- or DQ2.5-glia- $\omega$ 1-specific TCCs towards these two pairs of chimeras with identical 9-mer core sequence were similar, suggesting that flanking region is not contributing to the differential reactivity. We observed that both DQ2.5-glia- $\alpha$ 1a-specific- (Fig. 2C) and DQ2.5-glia- $\omega$ 1-specific TCCs (Fig. 2D) recognized chimeric p9 peptides suggesting that this residue is dispensable. The majority of specific TCCs lost their reactivity when the p7 residue was exchanged. However, these specific TCCs showed reactivity with chimeric peptides that had unchanged p7, i.e. DQ2.5-glia- $\alpha$ 1a- and DQ2.5-glia- $\omega$ 1-specific TCCs reacted with  $\omega$ 1\_Lp7Q and  $\alpha$ 1\_Qp7L, respectively. This observation highlights the importance of p7 in the specific T-cell recognition of DQ2.5-glia- $\alpha$ 1a or DQ2.5-glia- $\omega$ 1.

### High similarity in V-gene usage in TCRs specific for DQ2.5-glia- $\alpha$ 1a and DQ2.5-glia- $\omega$ 1

To examine the CeD patient T-cell repertoire reactive to HLA-DQ2.5:DQ2.5-glia- $\alpha$ 1a and HLA-DQ2.5:DQ2.5-glia- $\omega$ 1, we isolated single CD4<sup>+</sup> T cells directly from blood or gut biopsies of CeD patients using HLA-DQ2.5 tetramers and determined the sequences of rearranged TCR $\alpha$  and TCR $\beta$  gene pairs (Table 4). We looked at the overall usage of *TRAV* or *TRBV* gene segments.

In cases of dual V-gene usage, each V gene was given a half score. For DQ2.5-glia- $\alpha$ 1a, 165 single T cells isolated from 11 CeD patients produced 116 unique clonotypes expressing 37 *TRAV* and 26 *TRBV* genes (Fig. 3A and Fig. S1). Similar analysis of 113 DQ2.5-glia- $\omega$ 1-specific T cells isolated from 11 CeD patients produced 99 unique clonotypes expressing 35 *TRAV* and 30 *TRBV* genes (Fig. 3A and Fig. S1). We then calculated pair-wise Morisita-Horn index to analyze the similarity in *TRAV* and *TRBV* usage, where higher index value indicates higher similarity. We found high similarity in both *TRAV* and *TRBV* usage (0.74 and 0.65, respectively) between DQ2.5-glia- $\alpha$ 1a- and DQ2.5-glia- $\omega$ 1-specific TCRs (Fig. 3A, Fig. S2A). The similarity between DQ2.5-glia- $\alpha$ 1a- and DQ2.5-glia- $\omega$ 1-specific TCRs was highest when compared with the previously published (12) V-gene usage in T cells specific for DQ2.5-glia- $\alpha$ 2 and DQ2.5-glia- $\omega$ 2. (Fig. S2A).

We examined the five most expressed *TRAVs* and *TRBVs* in TCRs specific for DQ2.5-glia- $\alpha$ 1a- and DQ2.5-glia- $\omega$ 1 and observed similar genes being preferentially expressed (Fig. 3B). Of note, *TRBV29-1* and *TRBV20-1*, which were expressed as the most dominant *TRBV* in each of the two repertoires, are phylogenetically closely related (10). The *TRAV35*, *TRAV4*, *TRAV12-2*, *TRBV29-1* or *TRBV20-1* and *TRBV5-1* were the most frequently expressed *TRAV*- and *TRBV*-gene-segments in both TCR repertoires. Further, for both DQ2.5-glia- $\alpha$ 1a and DQ2.5-glia- $\omega$ 1-specific TCRs we did not observe preferential pairing of *TRAV* and *TRBV* (Fig. S3).

We also performed similar analysis of unique clonotypes obtained by paired TCR $\alpha$  and TCR $\beta$  sequencing of single CD4<sup>+</sup> T cells isolated from *in vitro* cultured T-cell line (TCL) generated from gut biopsies of CeD patients using the HLA-DQ2.5:DQ2.5-glia- $\alpha$ 1a ( $n = 50$ ) or HLA-DQ2.5:DQ2.5-glia- $\omega$ 1 tetramers ( $n = 20$ ) (Table 4 and Fig. S1, Fig. S2B). The *TRAV* and *TRBV* usage for both epitopes in data derived from T cells that were directly isolated from blood or gut biopsies was generally reproduced in the data obtained from TCLs.

We then aligned the CDR3 $\alpha$  and CDR3 $\beta$  sequences of the TCRs specific for DQ2.5-glia- $\alpha$ 1a and DQ2.5-glia- $\omega$ 1 (Fig. 3C) to analyze the CDR3 amino acid usage and positioning. The T cells were isolated using HLA-DQ2.5:DQ2.5-

glia- $\alpha$ 1a or HLA-DQ2.5:DQ2.5-glia- $\omega$ 1 tetramers, either directly from blood or gut biopsies or from *in vitro* cultured TCLs from gut biopsy. We did not observe any significant selection pattern in the amino acid usage and positioning in the CDR3 sequences.

We then examined the most frequently expressed V-genes in TCRs specific for the four immunodominant gluten epitopes (DQ2.5-glia- $\alpha$ 1a, DQ2.5-glia- $\omega$ 1, DQ2.5-glia- $\alpha$ 2 and DQ2.5-glia- $\omega$ 2). We analyzed unique clonotypes obtained by paired TCR $\alpha\beta$  sequencing of single T cells isolated either from TCLs or directly from blood or gut biopsies of CeD patients using HLA-DQ2.5 tetramers (Fig. S2B). *TRAV4* was among three most frequently used V-gene segments in the TCR repertoires specific for all the four epitopes. As all these gluten epitopes are restricted by HLA-DQ2.5, this feature of *TRAV4* bias could be dependent on HLA interactions.

### Crystal structures of the DQ2.5-glia- $\alpha$ 1a and DQ2.5-glia- $\omega$ 1 complexes

To gain an understanding of how two highly homologous epitopes give rise to two separate populations of T cells, we examined the corresponding peptide:HLA-II landscapes. Given the limited extent of cross-reactivity between the two gliadin determinants, it was unclear whether they would sit differently or within the same register in the HLA-DQ2.5 molecule. The observations made from the Ala substitution experiment performed on the TCCs is also based on the assumption that the substituted peptides bind in the identical binding registers. To establish this, we determined, to high resolution, the crystal structures of HLA-DQ2.5-glia- $\alpha$ 1a (2.0 Å) and HLA-DQ2.5-glia- $\omega$ 1 (2.6 Å) (Fig. 4A, 4B and Table 5) for data collection and refinement statistics). These complexes were solved with same linker peptide for both epitopes and accordingly any structural variation between these two binary complexes could be attributed to differences in the two epitopes bound within the antigen-binding cleft. While the crystal symmetry for the two crystals were different (P222 for DQ2.5-glia- $\alpha$ 1 versus C121 for DQ2.5-glia- $\omega$ 1), the crystal packing of the two structures was overall similar and, notably, with no crystal contacts involving the peptide positions p7 and p9. The HLA-DQ2.5:DQ2.5-glia- $\alpha$ 1a structure aligned well with previously determined binary HLA-DQ2.5:DQ2.5-glia- $\alpha$ 1a structure that was solved at lower resolution (16).

The higher resolution structure (reported here) will be compared to the HLA-DQ2.5:DQ2.5-glia- $\omega$ 1 structure. The electron densities corresponding to the two peptides were clear (Fig. S4).

Comparison of the HLA-DQ2.5:DQ2.5-glia- $\alpha$ 1a and the HLA-DQ2.5:DQ2.5-glia- $\omega$ 1 structures revealed minor differences (root mean square deviation (r.m.s.d.) = 0.32 Å for C $\alpha$  atoms of the peptide binding domain), revealing that the two epitopes did not bind in a different register within the HLA-DQ2.5 molecule (Fig. 4C, 4D). The p7 residue was an anchor residue in both complexes (solvent accessible surface area of p7L and p7Q was 29.5Å<sup>2</sup> and 20.7Å<sup>2</sup>, respectively, corresponding to 17% and 11% of the values for GXG). The interactions between the DQ2.5-glia- $\alpha$ 1a and DQ2.5-glia- $\omega$ 1 peptides and HLA-DQ2.5 were essentially conserved (Fig. 4A, 4B). Nevertheless, differences were apparent in the local environment of the peptide residues p7 and p9 (p7-Leu and p9-Tyr in HLA-DQ2.5:DQ2.5-glia- $\alpha$ 1a; p7-Gln and p9-Phe in HLA-DQ2.5:DQ2.5-glia- $\omega$ 1). These differences coincided with an adjustment (up to 0.9 Å) in the positioning of the HLA-DQ2.5  $\beta$  chain helix directly adjacent to p7 (Fig. 4C). The p7 difference between the peptides altered how each peptide was H-bonded to HLA-DQ2.5. Namely, in HLA-DQ2.5:DQ2.5-glia- $\alpha$ 1a, Lys71 $\beta$  formed a H-bond to the sidechain of p4-Gln, whereas in HLA-DQ2.5:DQ2.5-glia- $\omega$ 1, it H-bonded to the backbone of p5-Pro, thus freeing p4-Gln to form an additional H-bond to Ser28 $\beta$  at the floor of the peptide binding cleft (Fig. 4A, 4B). These structural perturbations also led to electrostatic alterations around the P7 residue (Fig. 4E, 4F). These differences are clearly sufficient to be discerned by T cells being specific for either of the two homologous epitopes.

To confirm that TCRs can discriminate between these two epitopes, we undertook surface plasmon resonance-based experiments using the HLA-DQ2.5-glia- $\alpha$ 1a responsive TCRs (S2 and L6) (10) run over HLA-DQ2.5-glia- $\alpha$ 1a and HLA-DQ2.5-glia- $\omega$ 1 (Fig. 5). We found that the S2 TCR specifically recognized HLA-DQ2.5-glia- $\alpha$ 1a only and hence discriminated between HLA-DQ2.5-glia- $\alpha$ 1 and HLA-DQ2.5-glia- $\omega$ 1. The L6 TCR, however, bound HLA-DQ2.5-glia- $\alpha$ 1a HLA-DQ2.5-glia- $\omega$ 1 with similar affinity. Taken together, the functional data and affinity

measurements suggest that, whilst a small population of TCRs is able to cross-react with the two highly homologous gliadin epitopes, the majority of TCRs lack cross-reactivity.

## Discussion

The CD4<sup>+</sup> T-cell response to gluten epitopes presented by the disease-predisposing HLA-DQ molecules is an essential part of the pathogenesis of CeD. Here we have studied two homologous and immunodominant gluten epitopes, DQ2.5-glia- $\alpha$ 1a (PFPQPELPY) and DQ2.5-glia- $\omega$ 1 (PFPQPEQPF). Despite only two amino acids difference in the 9-mer core region of the epitopes, the majority (~ 95 %) of T cells in blood and gut of CeD patients discriminate between the two epitopes with only a proportion of the T cells being cross-reactive. We found that the p7 residue was uniformly critical in both epitopes for discriminatory TCR recognition by all TCCs. Accordingly, we addressed the molecular basis underpinning the fine specificity of this response.

The p7 residues served as anchor residues in both HLA-DQ2.5:DQ2.5-glia- $\alpha$ 1a and HLA-DQ2.5:DQ2.5-glia- $\omega$ 1 and their sidechains were only partially exposed in the binary crystal structures. Notwithstanding, it is striking that the p7 residue is critical for specific T-cell recognition of both DQ2.5-glia- $\alpha$ 1a and DQ2.5-glia- $\omega$ 1 as well as DQ2.5-glia- $\alpha$ 2 and DQ2.5-glia- $\omega$ 2. In ternary crystal structures of three unique complexes of TCR:HLA-DQ2.5:DQ2.5-glia- $\alpha$ 2, the large sidechain of p7-Tyr forms critical interactions with the TCR, explaining its vital role for T-cell recognition of this epitope. However, the ternary structure of the S2 TCR (*TRAV4\*01-TRBV20-1\*01*) specific for the DQ2.5-glia- $\alpha$ 1a epitope in complex with HLA-DQ2.5 (10) revealed that this TCR does not make a direct contact with the less accessible p7-Leu residue. In spite of this, the TCC from which the S2 TCR was derived, along with ten other DQ2.5-glia- $\alpha$ 1a-specific TCCs, uniformly lost reactivity towards DQ2.5-glia- $\alpha$ 1a on p7 substitution in T-cell proliferation assays (10). The S2 TCR does discriminate between DQ2.5-glia- $\alpha$ 1a and DQ2.5-glia- $\omega$ 1, suggesting that even though the p7-Leu is not making direct contact to the TCR it is, in some way, influencing antigen recognition. To address this conundrum, we determined the crystal structures of HLA-DQ2.5 complexed with the DQ2.5-glia-

$\alpha$ 1a and DQ2.5-glia- $\omega$ 1 epitopes, which revealed that the two epitopes were accommodated by HLA-DQ2.5 using the same register. While the two peptide-HLA-DQ2.5 binary complexes were similar, differences around the p7 residue altered the neighboring sub-structure of the HLA-DQ2.5 molecule and associated electrostatic properties. Despite these relatively subtle differences in peptide:HLA-II topologies observed for the two immunodominant gluten epitopes, the S2 TCR must be able to detect them in a discriminatory manner. In the S2 TCR ternary complex, these p7 neighboring residues (specifically, HLA-DQ2.5 Asp66 $\beta$ , Glu69 $\beta$  and Arg70 $\beta$ ) form an extensive interface with CDR1 $\alpha$ , CDR3 $\beta$  and  $\alpha$ -framework residues of the S2 TCR, suggesting that the nature of this local environment is involved in epitope discrimination. Similar to the previous study, we found that for DQ2.5-glia- $\alpha$ 1a-specific TCCs the p7-Leu was uniformly critical in specific TCR recognition. The DQ2.5-glia- $\alpha$ 1a-specific TCCs analyzed in the two studies use many different *TRAV/TRBV* pairings suggesting that this critical role of p7 is not contingent of a particular TCR gene usage. Interestingly, the p7-Gln of the DQ2.5-glia- $\omega$ 1 epitope was also critical in specific TCR recognition for the DQ2.5-glia- $\omega$ 1-specific TCCs that also express fairly diverse TCRs. These observations suggest that the different p7 amino acids in DQ2.5-glia- $\alpha$ 1a and DQ2.5-glia- $\omega$ 1 epitopes induce subtle differences that are sensed by the TCRs giving discriminatory recognition of the two peptide:HLA-II complexes. Similar features of impact on TCR recognition by a MHC buried anchor residues, were observed for I-Ek and a hemoglobin epitope where substitution at the p6 residue affected T-cell recognition (17). Hence, these results indicate that anchor residues that are buried in the MHC II and make no direct contact with the TCR, can indirectly influence the specific TCR recognition.

The single-cell sequencing of TCRs specific for DQ2.5-glia- $\alpha$ 1a and DQ2.5-glia- $\omega$ 1 epitopes demonstrate the same V-gene bias. While V-gene bias is not a new phenomenon among TCRs specific for immunodominant gluten epitopes (6,8-12), the high similarity in V-gene usage and biased expression of common V genes between DQ2.5-glia- $\alpha$ 1a and DQ2.5-glia- $\omega$ 1 epitopes is noteworthy. However, the TCR bias is not as striking as in TCRs specific to DQ2.5-glia- $\alpha$ 2 where we observe biased expression of *TRBV7-2* paired with *TRAV26-1* and a conserved Arg in

CDR3 $\beta$  (8-10,12) or in TCRs specific for DQ8-glia- $\alpha$ 1 with biased expression of *TRAV26-2/TRBV9* and a conserved Arg in CDR3 $\alpha$  (6,11). In our hands, the TCR bias towards DQ2.5-glia- $\alpha$ 1a and DQ2.5-glia- $\omega$ 1 is neither characterized by preferential *TRAV/TRBV* pairing nor conservation of CDR3 motifs that are common for both epitopes. This variability in *TRAV/TRBV* pairing and CDR3 region could possibly result in variable TCR docking patterns.

Examining the wealth of TCR gene sequence data now accumulated for the four immunodominant HLA-DQ2.5-restricted gluten epitopes (DQ2.5-glia- $\alpha$ 1a and DQ2.5-glia- $\omega$ 1, DQ2.5-glia- $\alpha$ 2 and DQ2.5-glia- $\omega$ 2) in CeD (12), we observe that *TRAV4* is among the most frequently expressed *TRAVs* in T cells specific for all of the four epitopes. This *TRAV4* bias does not appear to be associated with a common *TRBV* usage or conservation of CDR3 sequences. The crystal structure of DQ2.5-glia- $\alpha$ 1a-specific TCR using *TRAV4* revealed that the *TRAV4* bias against DQ2.5-glia- $\alpha$ 1a is an effect of interactions between germline-encoded TCR residues, most prominently Tyr38 $\alpha$ , with residues of the  $\beta$ -chain of HLA-DQ2.5 (Arg70 $\beta$  and Arg77 $\beta$ ) (10). Similarly, HLA-DQ2.5:DQ2.5-glia- $\alpha$ 2 specific TCRs encoded by *TRAV26-1*, a phylogenetically close V gene with high sequence relatedness with *TRAV4*, formed analogous sets of interactions in ternary crystal structures. Therefore, we suggest that the common *TRAV4* bias in TCRs specific for immunodominant gluten epitopes restricted by HLA-DQ2.5 could be an outcome of conserved interaction between germline residues encoded by *TRAV4* and HLA-DQ2.5.

CeD provides an opportunity to study the natural immune response in humans towards a natural antigenic system that comprises a vast array of similar peptide sequences. Here we show that despite minor differences in peptide:HLA-II topologies and high similarity in TCR V-gene usage, the majority of the T cells discriminate between the two homologous and immunodominant gluten epitopes. For CeD, this implicates that highly homologous peptides have the potential to engage separate T-cell populations thereby resulting in broader and more robust T-cell responses to gluten. Of general implication, the study underscores the exquisite sensitivity of TCRs to detect subtle differences in peptide:HLA-II complexes.

## Material and Methods

### Patient material

We obtained ~60 ml of citrated full blood and six gut biopsies taken as part of a gastroduodenoscopy, using an Olympus H-180 endoscope and a regular biopsy forceps (Olympus), from both untreated and treated CeD patients. PBMCs were isolated by Ficoll-based density gradient centrifugation from the blood samples and cryopreserved for later use. Gut biopsies were processed to obtain single-cell suspension prior to cryopreservation. In brief, freshly collected biopsies were washed twice with 2 mM EDTA in 2% FCS at 37 °C for 10 min (to remove epithelial layer) followed by collagenase digestion (1 mg/ml) at 37 °C for 30-60 min, homogenization by syringe and filtration. We used previously established and cryopreserved T-cell lines (TCLs) that had been generated from single gut biopsy, of which some are published (18,19). Two new TCLs (CD1174 and CD1178) were generated by incubating gut biopsy with native chymotrypsin-digested gluten and deamidated chymotrypsin-digested gluten (both at 20  $\mu$ g/mL) for 3-5 days followed by addition of interleukin (IL)-2/IL-15.

### Tetramer staining, cell enrichment and FACS

Tetramers of recombinant HLA-DQ2.5 covalently linked with gluten-derived peptides containing the T-cell epitope DQ2.5-glia- $\alpha$ 1a peptide (QLQPFQPELPY, underlined 9-mer core amino acid sequence) and DQ2.5-glia- $\alpha$ 2 peptide (PQPELPYPQPE) were produced and conjugated with phycoerythrin-labeled streptavidin (Invitrogen) or allophycocyanin-labeled streptavidin (ProZyme) as described (20). Likewise, the HLA-DQ2.5:DQ2.5-glia- $\omega$ 1 (PQQPFQPEQFP) and HLA-DQ2.5:DQ2.5-glia- $\omega$ 2 (FPQPEQFPWQP) tetramers were generated with the same protocol. Tetramer staining was performed on cryopreserved PBMCs, on single-cell suspension prepared from gut biopsies and on TCLs. In brief, the cryopreserved PBMCs were thawed and stained with the fluorochrome-conjugated tetramers (10  $\mu$ g/ml each), for 30-45 min at room temperature followed by bead enrichment of tetramer-binding cells before adding surface antibody mix. Cryopreserved single-cell suspensions of gut biopsies were thawed and stained with fluorochrome-conjugated tetramers (10  $\mu$ g/ml) for 30-45 min at room temperature before adding surface antibody mix. Cryopreserved TCLs were

thawed and stained directly with fluorochrome-conjugated tetramers (10 µg/ml) for 2 hours at 37 °C before adding surface antibody mix. For PBMCs, cells within the singlet lymphocyte population were further gated to isolate tetramer-binding CD4<sup>+</sup> effector-memory T cells that were: CD3<sup>+</sup>, CD11c<sup>-</sup>, CD14<sup>-</sup>, CD15<sup>-</sup>, CD19<sup>-</sup>, CD56<sup>-</sup>, CD45RA<sup>-</sup>, CD62L<sup>-</sup>, and CD4<sup>+</sup>. For single-cell suspensions of gut biopsies, live cells within the singlet lymphocyte population were further gated to obtain tetramer-binding CD4<sup>+</sup> T cells that were: CD3<sup>+</sup>, CD11c<sup>-</sup>, CD14<sup>-</sup>, CD15<sup>-</sup>, CD19<sup>-</sup>, CD56<sup>-</sup>, CD8<sup>-</sup> and CD4<sup>+</sup>. For T-cell lines, live cells within the singlet lymphocyte population were further gated to obtain tetramer-binding CD4<sup>+</sup> T cells that were: CD3<sup>+</sup>, CD8<sup>-</sup> and CD4<sup>+</sup>. Finally, the populations of tetramer-binding CD4<sup>+</sup> T cells were index sorted for single-cell TCR sequencing or for generation of TCCs by in vitro expansion. The sorting was performed with a FACS Aria II instrument (BD Biosciences) at the Flow Cytometry Core Facility (Oslo University Hospital) and the flow cytometry data was analyzed with FlowJo software (FlowJo LLC). The antibodies used in the study were: CD62L-PerCP/Cy5.5, CD14-Pacific Blue, CD15-Pacific Blue, CD19-Pacific Blue and CD56-Pacific Blue; CD3-FITC, CD11c-Horizon V450, CD4-APC or CD4-APC-H7 (from BD Biosciences); CD45RA-PE-Cy7 and CD3-eVolve605, CD8-PE-Cy7 or CD8-PerCP (eBioscience). LIVE/DEAD marker fixable violet stain (Thermo Fischer Invitrogen) was used to exclude the dead cells.

### Generation and TCR sequencing of TCCs

The TCCs were generated from tetramer-sorted CD4<sup>+</sup> T cells and sequenced as described in previous publication (12). In brief, TCCs were generated using cloning by limited dilution and expanded without antigens followed by mRNA isolation, switching mechanism at the 5'-terminus of the RNA transcript based cDNA synthesis, PCR amplification and finally Sanger sequencing. TCCs from patient CD442 and CD1340 used for Ala scans in the current study were generated in the course of another project (7).

### T-cell proliferation assay

T-cell proliferation assays were carried out as described previously (21). In brief, 75,000 APCs (HLA-DQ2.5 homozygous Epstein-Barr virus (EBV)-transformed cells) were irradiated 75 Gy and incubated with 10 µM peptides at 37 °C for

24 h before the adding 50,000 T cells. 48 hours later, cultures were pulsed with 1 µCi <sup>3</sup>H-thymidine followed by harvesting and scintillation counting after 16-20 hours. The TCCs were identified as peptide-specific if the stimulation index (ratio of cpm after antigen stimulation and cpm after medium stimulation) was higher than 3.

### Single-cell TCR gene sequencing

We used a previously published protocol, based on nested PCR amplification using multiplex *TRAV* and *TRBV* primers, for single-cell TCR sequencing (22). The protocol was slightly modified by performing cDNA synthesis and the first PCR reaction in two separate steps. In short, single cells were sorted directly in 96-well plates containing 5 µl capture buffer (20 mM Tris-HCl (pH 8.0), 1% NP-40 and 1 U/µl RNase Inhibitor (optional)). The cDNA mix (5 µl of 1X SSII RT buffer, 1 mM dNTP, 2.5 mM DDT, 1 µM oligo d(T), 1 µM reverse *TRAC* (5'-AGTCAGATTTGTTGCTCCAGGCC-3') and *TRBC* (5'-TTCACCCACCAGCTCAGCTCC-3') primers, 1.5 U/µl RNase Inhibitor, 2.5 U/µl Superscript II in final 10 µl reaction volume) was added and cDNA synthesis was carried out at 42 °C for 50 min followed by an inactivation step at 72 °C for 10 min. The original protocol was followed to obtain a purified PCR product, which was then sequenced with 250-bp paired-end sequencing using Illumina MiSeq platform at the Norwegian Sequencing Centre (Oslo University Hospital). All the raw data generated in this study has been uploaded to the European Genome-phenome Archive (EGAS00001003245).

### Processing of TCR gene sequences

Processing of the raw sequencing data obtained from Illumina NGS was carried out in a multistep pipeline. The pipeline composed of: quality filtering of low-quality reads (Q < 30), annotation of the header with barcodes (row, plate and column) and gene (*TRA/TRB*) information, pairing and assembly of the annotated forward and reverse reads. These assembled reads were then collapsed to give up to three highest ranking reads (based on dupcount) to remove duplicates. In the next step V, D, J genes and CDR3 junction sequences were identified using IMGT-HighV-Quest online tool (23). A processing workflow implemented as an in-house Java program together with a custom MySQL database was used to further

extract the sequences. In brief, only productive sequences with dupcount > 100 were collapsed based on identical V gene, J gene and CDR3 (nucleotide level). These sequences were then re-filtered, so that only T cells with at least one TCR $\alpha$  and TCR $\beta$  chain and dual TCR $\alpha$  or TCR $\beta$  chains (but not both dual TCR $\alpha$  and TCR $\beta$ , maximum 3 chains) were considered for downstream analysis. We then categorized T cells as a clonotype if they have identical V and J gene (subgroup level), together with identical CDR3 region allowing for one nucleotide mismatch. This data was then used for further analysis of clonal expansion and V-gene usage. The paired TCR $\alpha$  and TCR $\beta$  sequencing of single T cells isolated from TCLs of CeD patients CD364, CD373, CD412 and CD436 used in V-gene usage analysis were generated in the course of another project for tracking clonotypes (7).

### **CDR3 amino acid sequence alignment**

CDR3 amino acid sequence alignment was performed using CLC sequence viewer v8.0 (CLC bio, Aarhus, Denmark) set to accurate with gap open penalty set to 20 and gap extension set to 4.

### **HLA-DQ2.5 production, crystallization and structure determination**

HLA-DQ2.5:DQ2.5-glia- $\alpha$ 1a-NC containing the epitope (QPFQPELPYP) and HLA-DQ2.5:DQ2.5-glia- $\omega$ 1 containing the epitope (QPFQPEQFPF) were expressed and purified as described previously (10). Briefly, Fos/Jun zipper fusion constructs of the HLA-DQ2.5  $\alpha$  and  $\beta$  chains with each peptide linked to the N-terminus of the HLA-DQ2.5  $\beta$  chain were co-expressed in Hi5 insect cells and purified via immobilized metal affinity and Superdex S200 gel filtration chromatography. Prior to crystallization, the Fos/Jun zippers were removed by enterokinase cleavage followed by Hitrap-Q ion exchange chromatography. HLA-DQ2.5:DQ2.5-glia- $\alpha$ 1a and HLA-DQ2.5:DQ2.5-glia- $\omega$ 1 were concentrated to 8 mg/mL and crystallized via hanging drop vapour diffusion at room temperature using mother liquor containing 23% PEG3350 and 0.1 M NaH<sub>2</sub>PO<sub>4</sub> (pH 6). Crystals were cryoprotected in mother liquor supplemented with 20% glycerol and were flash frozen in liquid N<sub>2</sub>. X ray diffraction data was collected at the MX1 and MX2 beamlines of the Australian synchrotron and processed using the programs XDS (24) and Scala (25) from the

CCP4 package (26). The structures of HLA-DQ2.5:DQ2.5-glia- $\alpha$ 1a and HLA-DQ2.5:DQ2.5-glia- $\omega$ 1 were solved via molecular replacement using Phaser (27), and contained two, and one protomers in the asymmetric unit, respectively. Structure models were generated by iterative rounds of model building in Coot (28) and restrained refinement in Phenix.

The coordinates and structure factors have been deposited in the Protein Data Bank under the following accession codes: HLA-DQ2.5:DQ2.5-glia- $\alpha$ 1a, 6MFG; HLA-DQ2.5:DQ2.5-glia- $\omega$ 1, 6MFF.

### **Surface plasmon resonance**

The surface plasmon resonance (SPR) experiments were conducted using Biacore 3000 (GE Healthcare) with Biacore CAPture sensor chip. This analysis was performed in order to obtain equilibrium affinity constants of TCRs toward HLA-DQ2.5:DQ2.5-glia- $\alpha$ 1a and HLA-DQ2.5:DQ2.5-glia- $\omega$ 1 constructs. TCRs were produced as described previously (10). The biotinylated ligands, HLA-DQ2.5:DQ2.5-glia- $\alpha$ 1a and HLA-DQ2.5:DQ2.5-glia- $\omega$ 1 and HLA-DQ2.5:DQ2.5-CLIP2 (ATPLLMQALPM, negative control for non-specific binding) were immobilized at 1000-1500 response units (RU) as previously shown (6,10). Flow cell surfaces were activated using 1:1 mixture of Biotin CAPture Reagent (Streptavidin and complementary oligonucleotide to the one on CAP surface) and HBS buffer (10 mM HEPES pH 7.4, 150 mM NaCl, 2 mM EDTA and 0.005% P20 detergent supplied by the manufacturer) at 25 °C using flow rate of 5  $\mu$ L/min. Following injection of 1 mM biotin, decreasing concentrations of TCR in HBS buffer were passed over the flow cells for 1.5 min time interval. The maximum concentration of TCR for dilution series was 200  $\mu$ M. The affinities of each TCR for HLA-DQ2.5:DQ2.5-glia- $\alpha$ 1a (LQPFQPELPY), HLA-DQ2.5:DQ2.5-glia- $\alpha$ 1a-NC (QPFQPELPYP) HLA-DQ2.5:DQ2.5-glia- $\omega$ 1 (QPFQPEQFPF) constructs were tested by conducting four, two and four independent experiments, respectively. The data for both HLA-DQ2.5-glia- $\alpha$ 1a constructs were combined since there was no measurable difference. The equilibrium dissociation constant, K<sub>D</sub>, was determined by fitting the data into non-linear one-site specific binding model using GraphPad Prism version 7.0.

### Statistical analysis

The Morisita-Horn index for analyzing similarity in *TRAV* and *TRBV* usage across TCRs specific for the four immunodominant epitopes was generated in R (3.3.2) with the use of the Vegan (2.4-2) package. We calculated pairwise similarities using `sim.table` function in Vegan package, setting 2 as the order of diversity measure (q) to generate the Morisita-Horn indices.

### Study approval

The study was approved by Regional Committee for Medical and Health Research Ethics South-East Norway (project 2010/2720). The patients participating in the study gave informed, written consent.

### Acknowledgments

We are grateful to the patients who participated in this study. We thank S. Furholm, M.H. Bakke and C. Hinrichs for collecting biological material from patients. We express our gratitude to B. Simonsen and S.R. Lund for producing the biotinylated HLA-DQ:gluten molecules and M.K. Johannesen for technical assistance.

### Funding sources

This study was supported by grants from Stiftelsen Kristian Gerhard Jebsen (project SKGJ-MED-017), the Research Council of Norway (project 179573/V40 through the Centre of Excellence funding scheme and project 233885) and the South-Eastern Norway Regional Health Authority (projects 2011050, 2013046, 2015009).

### Conflict of interest statement

The authors declare that they have no conflicts of interest with the contents of this article.

### Author contributions

S.D-K and L.C are joint first authors. S.D-K., L.C, J.P., L.F.R, R.S.N, A.C, H.H.R and S-W.Q. performed experiments and/or analyzed the data. K.E.L organized supply of CeD patient material. J.R and L.M.S are joint senior and corresponding authors, and wrote the manuscript together with S.D-K. All the authors critically revised and approved of the manuscript.

### References

1. Rossjohn, J., Gras, S., Miles, J. J., Turner, S. J., Godfrey, D. I., and

- McCluskey, J. (2015) T cell antigen receptor recognition of antigen-presenting molecules. *Annu. Rev. Immunol.* **33**, 169-200
2. Sollid, L. M., and Jabri, B. (2013) Triggers and drivers of autoimmunity: lessons from coeliac disease. *Nat.Rev.Immunol.* **13**, 294-302
3. Bromilow, S., Gethings, L. A., Buckley, M., Bromley, M., Shewry, P. R., Langridge, J. I., and Clare Mills, E. N. (2017) A curated gluten protein sequence database to support development of proteomics methods for determination of gluten in gluten-free foods. *J. Proteomics* **163**, 67-75
4. Sollid, L. M., Qiao, S. W., Anderson, R. P., Gianfrani, C., and Koning, F. (2012) Nomenclature and listing of celiac disease relevant gluten T-cell epitopes restricted by HLA-DQ molecules. *Immunogenetics* **64**, 455-460
5. Tye-Din, J. A., Stewart, J. A., Dromey, J. A., Beissbarth, T., van Heel, D. A., Tatham, A., Henderson, K., Mannering, S. I., Gianfrani, C., Jewell, D. P., Hill, A. V., McCluskey, J., Rossjohn, J., and Anderson, R. P. (2010) Comprehensive, quantitative mapping of T cell epitopes in gluten in celiac disease. *Sci.Transl.Med.* **2**, 41ra51
6. Broughton, S. E., Petersen, J., Theodossis, A., Scally, S. W., Loh, K. L., Thompson, A., van, B. J., Kooy-Winkelaar, Y., Henderson, K. N., Beddoe, T., Tye-Din, J. A., Mannering, S. I., Purcell, A. W., McCluskey, J., Anderson, R. P., Koning, F., Reid, H. H., and Rossjohn, J. (2012) Biased T cell receptor usage directed against human leukocyte antigen DQ8-restricted gliadin peptides is associated with celiac disease. *Immunity.* **37**, 611-621
7. Risnes, L. F., Christophersen, A., Dahal-Koirala, S., Neumann, R. S., Sandve, G. K., Sarna, V. K., Lundin,

- K. E., Qiao, S. W., and Sollid, L. M. (2018) Disease-driving CD4<sup>+</sup> T cell clonotypes persist for decades in celiac disease. *J. Clin. Invest.* **128**, 2642-2650
8. Qiao, S. W., Raki, M., Gunnarsen, K. S., Loset, G. A., Lundin, K. E., Sandlie, I., and Sollid, L. M. (2011) Posttranslational modification of gluten shapes TCR usage in celiac disease. *J. Immunol.* **187**, 3064-3071
  9. Qiao, S. W., Christophersen, A., Lundin, K. E., and Sollid, L. M. (2014) Biased usage and preferred pairing of  $\alpha$ - and  $\beta$ -chains of TCRs specific for an immunodominant gluten epitope in coeliac disease. *Int. Immunol.* **26**, 13-19
  10. Petersen, J., Montserrat, V., Mujico, J. R., Loh, K. L., Beringer, D. X., van, L. M., Thompson, A., Mearin, M. L., Schweizer, J., Kooy-Winkelaar, Y., van, B. J., Drijfhout, J. W., Kan, W. T., La Gruta, N. L., Anderson, R. P., Reid, H. H., Koning, F., and Rossjohn, J. (2014) T-cell receptor recognition of HLA-DQ2-gliadin complexes associated with celiac disease. *Nat. Struct. Mol. Biol.* **21**, 480-488
  11. Petersen, J., van Bergen, J., Loh, K. L., Kooy-Winkelaar, Y., Beringer, D. X., Thompson, A., Bakker, S. F., Mulder, C. J., Ladell, K., McLaren, J. E., Price, D. A., Rossjohn, J., Reid, H. H., and Koning, F. (2015) Determinants of gliadin-specific T cell selection in celiac disease. *J. Immunol.* **194**, 6112-6122
  12. Dahal-Koirala, S., Risnes, L. F., Christophersen, A., Sarna, V. K., Lundin, K. E., Sollid, L. M., and Qiao, S. W. (2016) TCR sequencing of single cells reactive to DQ2.5-glia- $\alpha$ 2 and DQ2.5-glia- $\omega$ 2 reveals clonal expansion and epitope-specific V-gene usage. *Mucosal Immunol.* **9**, 587-596
  13. Petersen, J., Kooy-Winkelaar, Y., Loh, K. L., Tran, M., van Bergen, J., Koning, F., Rossjohn, J., and Reid, H. H. (2016) Diverse T cell receptor gene usage in HLA-DQ8-associated celiac disease converges into a consensus binding solution. *Structure* **24**, 1643-1657
  14. Christophersen, A., Risnes, L. F., Bergseng, E., Lundin, K. E., Sollid, L. M., and Qiao, S. W. (2016) Healthy HLA-DQ2.5<sup>+</sup> Subjects Lack Regulatory and Memory T Cells Specific for Immunodominant Gluten Epitopes of Celiac Disease. *J. Immunol.* **196**, 2819-2826
  15. Sarna, V. K., Lundin, K. E. A., Morkrid, L., Qiao, S. W., Sollid, L. M., and Christophersen, A. (2018) HLA-DQ-Gluten Tetramer Blood Test Accurately Identifies Patients With and Without Celiac Disease in Absence of Gluten Consumption. *Gastroenterology* **154**, 886-896.e886
  16. Kim, C. Y., Quarsten, H., Bergseng, E., Khosla, C., and Sollid, L. M. (2004) Structural basis for HLA-DQ2-mediated presentation of gluten epitopes in celiac disease. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 4175-4179
  17. Kersh, G. J., Miley, M. J., Nelson, C. A., Grakoui, A., Horvath, S., Donermeyer, D. L., Kappler, J., Allen, P. M., and Fremont, D. H. (2001) Structural and functional consequences of altering a peptide MHC anchor residue. *J. Immunol.* **166**, 3345-3354
  18. Molberg, O., McAdam, S. N., Korner, R., Quarsten, H., Kristiansen, C., Madsen, L., Fugger, L., Scott, H., Noren, O., Roepstorff, P., Lundin, K. E., Sjostrom, H., and Sollid, L. M. (1998) Tissue transglutaminase selectively modifies gliadin peptides that are recognized by gut-derived T cells in celiac disease. *Nat. Med.* **4**, 713-717
  19. Arentz-Hansen, H., Korner, R., Molberg, O., Quarsten, H., Vader, W., Kooy, Y. M., Lundin, K. E., Koning, F., Roepstorff, P., Sollid, L. M., and

- McAdam, S. N. (2000) The intestinal T cell response to  $\alpha$ -gliadin in adult celiac disease is focused on a single deamidated glutamine targeted by tissue transglutaminase. *J. Exp. Med.* **191**, 603-612
20. Quarsten, H., McAdam, S. N., Jensen, T., Arentz-Hansen, H., Molberg, O., Lundin, K. E., and Sollid, L. M. (2001) Staining of celiac disease-relevant T cells by peptide-DQ2 multimers. *J. Immunol.* **167**, 4861-4868
21. Christophersen, A., Raki, M., Bergseng, E., Lundin, K. E., Jahnsen, J., Sollid, L. M., and Qiao, S. W. (2014) Tetramer-visualized gluten-specific CD4<sup>+</sup> T cells in blood as a potential diagnostic marker for coeliac disease without oral gluten challenge. *United European Gastroenterol J* **2**, 268-278
22. Han, A., Glanville, J., Hansmann, L., and Davis, M. M. (2014) Linking T-cell receptor sequence to functional phenotype at the single-cell level. *Nat. Biotechnol.* **32**, 684-692
23. Lefranc, M. P., Giudicelli, V., Ginestoux, C., Bodmer, J., Muller, W., Bontrop, R., Lemaitre, M., Malik, A., Barbie, V., and Chaume, D. (1999) IMGT, the international ImMunoGeneTics database. *Nucleic Acids Res.* **27**, 209-212
24. Kabsch, W. (2010) Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 133-144
25. Evans, P. (2006) Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 72-82
26. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235-242
27. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674
28. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486-501

### Figure legends

**FIGURE 1. DQ2.5-glia- $\alpha$ 1a- and DQ2.5-glia- $\omega$ 1-specific CD4<sup>+</sup> T cells in peripheral blood and gut of CeD patients.** (A) Representative plot showing tetramer staining of CD4<sup>+</sup> T cells with HLA:DQ gluten tetramers DQ2.5:DQ2.5-glia- $\alpha$ 1a and DQ2.5:DQ2.5-glia- $\omega$ 1 in blood and gut of CeD patients. The peripheral blood mononuclear cells (PBMCs) were stained with tetramers followed by bead enrichment of the tetramer-positive cells. The flow plot shows tetramer-binding CD4<sup>+</sup> effector-memory T-cells. \*Frequency of tetramer-positive CD4<sup>+</sup> T-cells per million total CD4<sup>+</sup> T-cells. The gut biopsies were stained with HLA:DQ gluten tetramers, however bead enrichment was not performed. The plot in the figure shows tetramer-binding CD4<sup>+</sup> T cells (B) Relative ratio of CD4<sup>+</sup> T cells positive to both DQ2.5:DQ2.5-glia- $\alpha$ 1a- and DQ2.5:DQ2.5-glia- $\omega$ 1-tetramers with CD4<sup>+</sup> T cells that are positive to either of the two tetramers in samples obtained from 11 CeD patients. The horizontal line represents the mean value. (C) Staining of TCCs generated from HLA-DQ2.5:DQ2.5-glia- $\alpha$ 1a or HLA-DQ2.5:DQ2.5-glia- $\omega$ 1 tetramer binding CD4<sup>+</sup> T cells isolated from peripheral blood with HLA-DQ2.5:gluten tetramers (DQ2.5-glia- $\alpha$ 1a, DQ2.5-glia- $\alpha$ 2, DQ2.5-glia- $\omega$ 1, DQ2.5-glia- $\omega$ 2) and CLIP2 tetramer. TCC1383P.A.2 is not included in the figure due to lack of TCC to carry out this staining (all tetramers attached with PE). However, TCC1383P.A.2 was tested in another staining experiment (tetramers attached with APC or PE) where it stained distinctly only with DQ2.5-glia- $\alpha$ 1a (Data not

shown). **(D)** T-cell proliferation of the TCCs generated from HLA-DQ2.5:DQ2.5-glia- $\alpha$ 1a or HLA-DQ2.5:DQ2.5-glia- $\omega$ 1 tetramer binding CD4+ T cells with panel peptides representing the known HLA-DQ2.5 restricted epitopes. DQ2.5-glia- $\alpha$ 1a (QLQPFQPELPY, underlined 9-mer core amino acid sequence), DQ2.5-glia- $\alpha$ 1b (PQPELPYQPELPY), DQ2.5-glia- $\alpha$ 2 (PQPELPYQPPQL), DQ2.5-glia- $\gamma$ 1 (PEQPQSFPEQERP), DQ2.5-glia- $\gamma$ 2 (GIHQEQPAQL), DQ2.5-glia- $\gamma$ 3 (FPQQPEQYPQQP), DQ2.5-glia- $\gamma$ 4a (FSQPEQEFPQPQ), DQ2.5-glia- $\gamma$ 4b (FPQPEQEFPQPQ), DQ2.5-glia- $\gamma$ 4c (TEQPEQFPFPQ), DQ2.5-glia- $\gamma$ 5 (PEQPFPEQPEQ), DQ2.5-glia- $\omega$ 1 (PQQPFQPEQPFQ), DQ2.5-glia- $\omega$ 2 (FPQPEQFPWQP), DQ2.5-hor3 (PIPEQPQYPQ) and DQ2.5-ave1a (YQYPQEPEPFV). Two independent experiments were performed with measurements performed either in duplicate or triplicate. The data shows a representative plot where error bars represent mean $\pm$ s.d for the triplicates.

**FIGURE 2. Effect of amino acid substitutions in the DQ2.5-glia- $\alpha$ 1a and DQ2.5-glia- $\omega$ 1 epitopes on T-cell recognition.** Reactivity patterns of (A) DQ2.5-glia- $\alpha$ 1a-specific- and (B) DQ2.5-glia- $\omega$ 1-specific TCCs to the single Ala-substituted DQ2.5-glia- $\alpha$ 1a and DQ2.5-glia- $\omega$ 1 peptides normalized to the wild-type response (1) are shown. The reactivity of (C) DQ2.5-glia- $\alpha$ 1a-specific- and (D) DQ2.5-glia- $\omega$ 1-specific TCCs against chimeras:  $\alpha$ 1a\_Qp7L,  $\alpha$ 1a\_Fp9Y,  $\omega$ 1\_Lp7Q and  $\omega$ 1\_Yp9F peptides normalized to the wild-type response (1) are shown. The data shown is average of three independent experiments each with measurements performed in triplicates. Error bars represent mean $\pm$ s.d for three experiments.

**FIGURE 3. DQ2.5-glia- $\alpha$ 1a- and DQ2.5-glia- $\omega$ 1-specific TCRs.** Unique T-cell clonotypes obtained from sequencing of TCRs (pairs of rearranged TCR $\alpha$  and TCR $\beta$  genes) of DQ2.5-glia- $\alpha$ 1a and DQ2.5-glia- $\omega$ 1-specific T cells isolated directly from blood or gut biopsies of CeD patients using corresponding HLA:DQ tetramers were analyzed **(A and B)**. Panel **A** shows frequencies of *TRAV* and *TRBV* usage and Morisita-Horn index for pair-wise comparison of similarity in *TRAV* and *TRBV* usage. Panel **B** shows frequency of the five most dominant

*TRAV*s (upper panel) and *TRBV*s (lower panel) among the unique clonotypes observed in the respective TCR repertoire. The total number of clonotypes analyzed in each sample is denoted in X axis. CDR3 $\alpha$  and CDR3 $\beta$  amino acid sequences of the DQ2.5-glia- $\alpha$ 1a and DQ2.5-glia- $\omega$ 1-specific T cells isolated either directly from blood or gut biopsies or from in vitro cultured TCLs from gut biopsy were aligned and analyzed **(C)** for the usage of amino acids at different positions.

**FIGURE 4. Structures of HLA-DQ2.5:HLA-DQ2.5-glia- $\alpha$ 1a and HLA-DQ2.5:HLA-DQ2.5-glia- $\omega$ 1 complexes.** HLA-DQ2.5 shown as cartoon and peptides as sticks, with HLA-DQ2.5  $\alpha$ -chain in light blue,  $\beta$ -chain in yellow and DQ2.5-glia- $\alpha$ 1a and DQ2.5-glia- $\omega$ 1 peptides in grey green, respectively. **(A)** Interactions between HLA-DQ2.5 and the DQ2.5-glia- $\alpha$ 1a peptide **(B)** Interactions between HLA-DQ2.5 and the DQ2.5-glia- $\omega$ 1 peptide. **(C)** Overlay of HLA-DQ2.5:DQ2.5-glia- $\alpha$ 1a and HLA-DQ2.5:DQ2.5-glia- $\omega$ 1 (top view) showing differences in the HLA-DQ2.5  $\beta$ -chain helix. The pointers show the regions where the helix is shifted by  $>0.5\text{\AA}$  **(D)** Overlay of HLA-DQ2.5:DQ2.5-glia- $\alpha$ 1a and HLA-DQ2.5:DQ2.5-glia- $\omega$ 1 (side view) showing differences in sidechain conformation in p9. **(E)** Electrostatic surface potential of HLA-DQ2.5:DQ2.5-glia- $\alpha$ 1a, and **(F)** of HLA-DQ2.5:DQ2.5-glia- $\omega$ 1.

**FIGURE 5. Surface plasmon resonance analysis of TCR interactions with HLA-DQ2.5-glia- $\alpha$ 1a and HLA-DQ2.5-glia- $\omega$ 1.** Sensorgram and binding curve fit of S2 **(A, B)** and L6 **(C, D)** affinity data into one-site specific binding model toward **(A, C)** HLA-DQ2.5-glia- $\alpha$ 1a and **(B, D)** HLA-DQ2.5-glia- $\omega$ 1. For  $K_D$  determination, all SPR data was derived from six (n=6) and four (n=4) independent experiments for HLA-DQ2.5-glia- $\alpha$ 1a and HLA-DQ2.5-glia- $\omega$ 1, respectively. SPR, surface plasmon resonance. Req, Response at equilibrium. RU, response units.  $K_D$ , dissociation constant. Error bars, s.d.