

Structural insights into clonal restriction and diversity in T cell recognition of two immunodominant SARS-CoV-2 nucleocapsid epitopes

Received: 14 February 2025

Accepted: 5 November 2025

Published online: 10 December 2025

 Check for updates

Ping Yuan^{1,9}, Guodong Chen^{1,9}, Yukun Li^{2,9}, Xichun Liu³,
Shayana Saravanakumar^{4,5}, Jianfeng Zhao¹, Qianyu Ji⁶, Hong Wang⁷,
Ying-Wu Lin³, Mostafa Elbahnasawy⁸, Nan-Ping Weng⁸,
Brian G. Pierce^{4,5} ✉, Roy A. Mariuzza^{4,5} ✉ & Daichao Wu¹ ✉

T cells play a crucial role in clearing SARS-CoV-2 and in forming long-term memory responses to that coronavirus. The highly immunogenic nucleocapsid (N) protein of SARS-CoV-2 is much more conserved than the spike (S) protein across variants of concern, making it an attractive vaccine target for activating cytotoxic CD8⁺ T cells. Of particular interest are the immunodominant N epitopes LLL and SPR. Whereas LLL elicits a clonally restricted T cell response, the response to SPR is highly diverse. To understand the basis for this difference, here we determine structures of T cell receptors (TCRs) bound to LLL–HLA-A2 and SPR–HLA-B7, revealing the structural underpinnings of highly restricted V α gene usage by LLL-specific TCRs, as well as multiple structural solutions to recognizing SPR and thereby generating a clonally diverse T cell response to that epitope. These structures also provide frameworks for understanding T cell recognition of SARS-CoV-2 variants and other coronaviruses. Finally, we compare the X-ray structures of TCR–LLL–HLA-A2 and TCR–SPR–HLA-B7 complexes with models predicted by multiple versions of AlphaFold, highlighting some success while showing room for improvement. Overall, our findings expand understanding of coronavirus T cell recognition, informing vaccine design and advances in computational modeling approaches.

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) caused the global coronavirus disease 2019 (COVID-19) pandemic^{1–3}. Understanding the mechanisms governing the adaptive immune response to SARS-CoV-2 is vital for predicting the outcome of infections and evaluating vaccine efficacy. Neutralizing antibodies, CD4⁺ helper T cells, and CD8⁺ killer T cells all contribute to the control of SARS-CoV-2 and the protection offered by vaccines, although their relative importance in the context of infection and prophylaxis remains to be elucidated^{4–6}. Neutralizing antibodies are clearly protective⁷, but may be short-lived

and are not elicited in all infected individuals^{6,8}. CD8⁺ T cells play a crucial role in clearing SARS-CoV-2 and forming long-term memory responses to this coronavirus^{9–12}. Indeed, there are numerous cases of healthy individuals successfully controlling SARS-CoV-2 infection in the absence of detectable neutralizing antibodies but with prominent SARS-CoV-2-specific T cell memory^{5,13–17}. In contrast to epitopes recognized by antibodies, which are sensitive to mutations causing viral escape, CD8⁺ T cells recognize epitopes from both variable and highly conserved viral proteins, thus offering longer immune protection^{18,19}.

A full list of affiliations appears at the end of the paper. ✉ e-mail: pierce@umd.edu; rmariuzz@umd.edu; wudaichao@usc.edu.cn

Extensive studies have identified SARS-CoV-2 epitopes that elicit protective T cell responses to this virus and also delineated T cell repertoires specific for these epitopes²⁰. T cell responses have been detected to multiple open reading frames encoding both structural (S, M, N) and nonstructural (nsp3, 4, 6, 7, 12, and 13) proteins, with S (spike) and N (nucleocapsid) proteins eliciting the most robust CD8⁺ and CD4⁺ T cell responses. The N protein, which functions in viral RNA genome packaging and assembly into particles, is abundant and much more conserved than the S protein across SARS-CoV-2 variants of concern (VOCs)^{21,22}. This sequence conservation makes the highly immunogenic N protein an attractive vaccine target for activating cytotoxic CD8⁺ T cells. The N protein triggers both antibody and T cell responses that correlate with control of SARS-CoV-2 infection in humans and the K18-hACE2 mouse model^{23–26}. Longitudinal studies of SARS-CoV-1-recovered patients have shown that N-specific memory T cells are sustained for up to 11 years¹⁹ to 17 years²⁷, suggesting that N-specific cellular immunity against SARS-CoV-2 may also be long-lasting.

Of particular interest among SARS-CoV-2 N protein epitopes are N_{222–230} (LLDRNQL, designated LLL), which is presented by HLA-A*02:01²⁸, and N_{105–113} (SPRWYFYLL, designated SPR), which is presented by HLA-B*07:02^{29–31}. Both epitopes are immunodominant. SARS-CoV-2 infection was found to establish a stable and age-independent response against LLL²⁸. In addition, LLL is one of six SARS-CoV-2 T cell epitopes included in a peptide-based vaccine against COVID-19 (CoVac-1)^{32,33}. This vaccine induced T cell responses in a Phase I/II clinical that were unaffected by current VOCs. The LLL peptide is also a component of a T cell-directed mRNA vaccine (BNT162b4) that protected hamsters against severe disease³⁴. A striking feature of the T cell response to the LLL epitope is a lack of sequence diversity: >50% of HLA-A*02:01-restricted, LLL-specific TCRs from COVID-19 convalescent patients (CPs) used the nearly identical TRAV12-1 or TRAV12-2 gene segments with limited CDR3α motifs²⁸.

T cell responses against SPR–HLA-B*07:02 are among the most dominant identified to date in SARS-CoV-2-infected individuals^{16,30,35,36}. SPR-specific T cells are associated with less severe COVID-19 disease and high antiviral efficacy³¹. These T cells were maintained 6 months after infection with preserved activity against Alpha, Beta, Gamma, and Delta SARS-CoV-2 variants, suggesting durable protective immunity. Furthermore, CD8⁺ T specific for SPR–HLA-B*07:02 were detected at high frequencies in pre-pandemic samples and displayed cross-reactivity toward circulating OC43 and HKU-1 betacoronaviruses^{29,30}. In sharp contrast to LLL–HLA-A*02:01-specific TCRs²⁸, TCRs from COVID-19 CPs specific for the HLA-B*07:02-restricted SPR epitope were highly diverse and utilized a wide variety of unrelated α/β chain pairs, including TRAV25/TRBV7-8, TRAV17/TRBV6-6, TRAV13-1/TRBV29-1, and TRAV4/TRBV7-3³⁰. Such diversity in antiviral T cell responses is believed to provide T cell functional heterogeneity and assure protection against viral escape³⁷.

With only one exception²⁸, that of TCR LLL8 bound to LLL and HLA-A2, previous structural studies of TCR recognition of SARS-CoV-2 have been confined to epitopes derived from the S protein^{38–42}. To advance our understanding of TCR recognition N epitopes, we determined crystal structures of a second LLL-specific TCR (LLL6E) bound LLL–HLA-A2 and of two SPR-specific TCRs (Q04 and CLB) bound to SPR–HLA-B7. The LLL6E–LLL–HLA-A2 complex revealed the basis for dominant usage of TRAV12-1 and TRAV12-2 gene segments by LLL-specific TCRs and for the ability of α chains encoded by these genes to pair with diverse β chains. The Q04–SPR–HLA-B7 and CLB–SPR–HLA-B7 complexes demonstrated that there are multiple structural solutions to recognizing SPR. This clonally diverse T cell response may help prevent viral escape though epitope mutations. Furthermore, structures of TCRs bound to SPR–HLA-B7 and LLL–HLA-A2 provide a framework for understanding T cell recognition of SARS-CoV-2 variants

and homologous N epitopes from other human coronaviruses at the atomic level. Finally, we compared the X-ray structures of the LLL6E–LLL–HLA-A2, Q04–SPR–HLA-B7, and CLB–HLA-B7 complexes with models predicted by the deep learning method AlphaFold⁴³. The results provide valuable insights into the accuracy and limitations of AlphaFold in modeling TCR–pMHC complexes.

Results

Interaction of SARS-CoV-2-specific TCRs with nucleocapsid epitopes SPR and LLL

TCRs Q004 and Clone B (referred to here as Q04 and CLB, respectively) were isolated by screening CD8⁺ T cells from COVID-19 CPs with SPR–HLA-B7 tetramers^{29,30}. Q04 and CLB were the dominant clonotypes in patients Q004 and CA13, respectively. These TCRs use completely different α/β chain pairs. Q04 utilizes gene segments TRAV25 and TRAJ40 for the α chain, and TRBV7-8 and TRBJ1-2 for the β chain, whereas CLB utilizes TRAV17 and TRAJ57 for the α chain, and TRBV6-6 and TRBJ2-7 for the β chain (Supplementary Table 1).

TCR LLL6 was isolated from a COVID-19 CP using LLL–HLA-A2 tetramers²⁸. LLL6 uses TRAV12-2 and TRAJ17 for the α chain, and TRBV9 and TRBJ2-7 for the β chain (Supplementary Table 1). We also engineered a variant of LLL6 (designated LLL6E) in which TRAV12-2 was replaced by TRAV12-1 without altering the sequence of complementarity-determining region 3α (CDR3α): ⁸⁸CVQGAAGNKLTF⁹⁹. However, TRAV12-1 and TRAV12-2 have different but closely related CDR1α and CDR2α sequences: ²⁷NSASQS³² and ²⁷DRGSQS³² for TRAV12-1 and TRAV12-2, respectively, and ⁵⁰VYSSGN⁵⁵ and ⁵⁰IYSNGD⁵⁵ for TRAV12-1 and TRAV12-2, respectively. The LLL6E variant was used for crystallizing a complex with LLL–HLA-A2 because wild-type LLL6 did not co-crystallize with this ligand, for unknown reasons. Of note, both TRAV12-1 and TRAV12-2 are used by LLL-specific TCRs, although TRAV12-2 is more prevalent (50% of 6,695 unique TCR sequences versus 6% for TRAV12-1)²⁸.

We used surface plasmon resonance to measure the affinity of TCRs Q04, CLB, LLL6, and LLL6E for HLA-B7 loaded with SPR peptide or HLA-A2 loaded with LLL peptide (Fig. 1). Recombinant TCR and pMHC proteins were expressed by *in vitro* folding from *E. coli* inclusion bodies. Biotinylated SPR–HLA-B7 or LLL–HLA-A2 was directionally coupled to a biosensor surface and increasing concentrations of TCR were flowed sequentially over the immobilized pMHC ligand. Q04 and CLB bound SPR–HLA-B7 with dissociation constants (K_D s) of 0.43 μM and 0.41 μM, respectively (Fig. 1a, b). Kinetic parameters (on- and off-rates) for the binding of TCR Q04 to SPR–HLA-B7 were $k_{on} = 1.7 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$ and $k_{off} = 0.068 \text{ s}^{-1}$, corresponding to a K_D of 0.40 μM (Fig. 1a), which matches the K_D from equilibrium analysis. For TCR CLB, kinetic parameters were $k_{on} = 2.1 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$ and $k_{off} = 0.076 \text{ s}^{-1}$, corresponding to a K_D of 0.36 μM (Fig. 1b), which is similar to the K_D from equilibrium analysis. LLL6 bound LLL–HLA-A2 with a K_D of 3.6 μM, with $k_{on} = 2.9 \times 10^4 \text{ M}^{-1} \text{ s}^{-1}$ and $k_{off} = 0.081 \text{ s}^{-1}$, corresponding to a K_D of 2.8 μM (Fig. 1c). For TCR LLL6E, a K_D of 15.2 μM was obtained under equilibrium conditions (Fig. 1d). As this K_D is only fourfold higher than that of wild-type LLL6 (3.6 μM), replacement of TRAV12-1 by TRAV12-2 did not have a major impact on affinity, despite several amino acid differences in CDR1α and CDR2α.

Overview of the TCR–SPR–HLA-B7 and TCR–LLL–HLA-A2 complexes

To understand how TCRs Q04, CLB, and LLL6E recognize their cognate nucleocapsid epitopes and to explain the effect of sequence differences or mutations in these epitopes on recognition, we determined the structures of the Q04–SPR–HLA-B7, CLB–SPR–HLA-B7, and LLL6E–LLL–HLA-A2 complexes at 2.75, 2.04, and 2.17 Å resolution, respectively (Supplementary Table 2 and Fig. 2a–c). The interface between TCR and pMHC is in unambiguous electron density in all complex structures (Supplementary Fig. 1). The Q04–SPR–HLA-B7

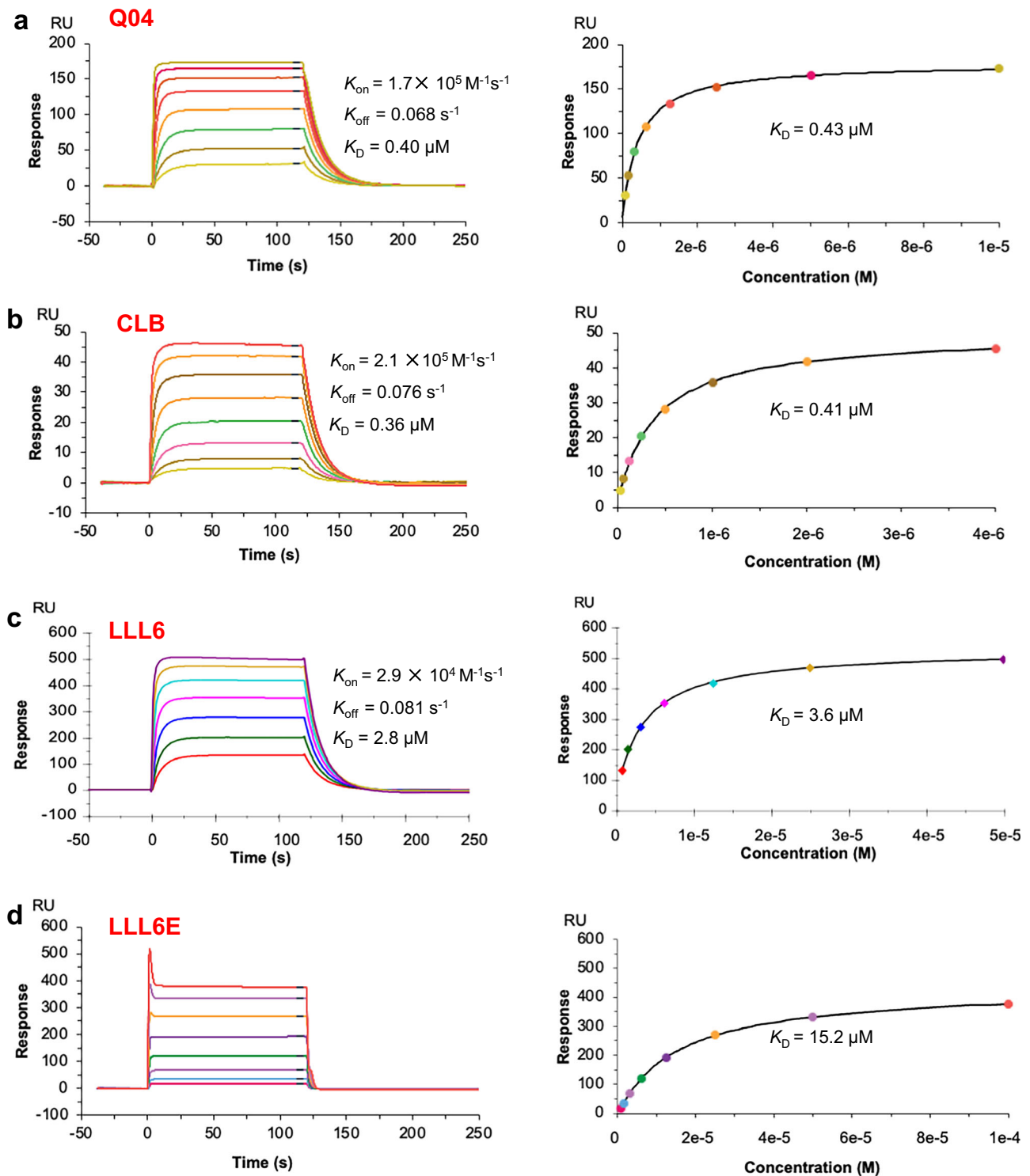


Fig. 1 | Surface plasmon resonance analysis of SARS-CoV-2-specific TCRs binding to nucleocapsid epitopes. **a** (left) TCR Q04 at concentrations of 0.08, 0.16, 0.31, 0.63, 1.25, 2.5, 5, and 10 μM was injected over immobilized SPR-HLA-B7 (300 RU). (right) Fitting curve for equilibrium binding that resulted in a K_D of 0.43 μM . **b** (left) TCR CLB at concentrations of 0.03, 0.06, 0.13, 0.25, 0.5, 1, 2, and 4 μM was injected over immobilized SPR-HLA-B7 (200 RU). (right) Fitting curve for equilibrium binding that resulted in a K_D of 0.41 μM . **c** (left) Wild-type LLL6 at

concentrations of 0.78, 1.56, 3.12, 6.3, 12.5, 25, and 50 μM was injected over immobilized LLL-HLA-A2 (1000 RU). (right) Fitting curve for equilibrium binding that resulted in a K_D of 3.6 μM . **d** (left) TCR LLL6E at concentrations of 0.78, 1.56, 3.12, 6.25, 12.5, 25, 50, and 100 μM was injected over immobilized LLL-HLA-A2 (1000 RU). (right) Fitting curve for equilibrium binding that resulted in a K_D of 15.2 μM .

crystal contains four complex molecules in the asymmetric unit. The root-mean-square difference (RMSD) in α -carbon positions for the TCR $V\alpha V\beta$ and MHC $\alpha 1\alpha 2$ modules, including the SPR peptide, is 0.30–0.44 \AA for the four Q04–SPR–HLA-B7 complexes. Based on this

close similarity, the following description of Q04–SPR–HLA-B7 interactions applies to all molecules in the asymmetric unit.

Both Q04 and CLB dock symmetrically over SPR–HLA-B7 in a canonical diagonal orientation, but with moderately different crossing

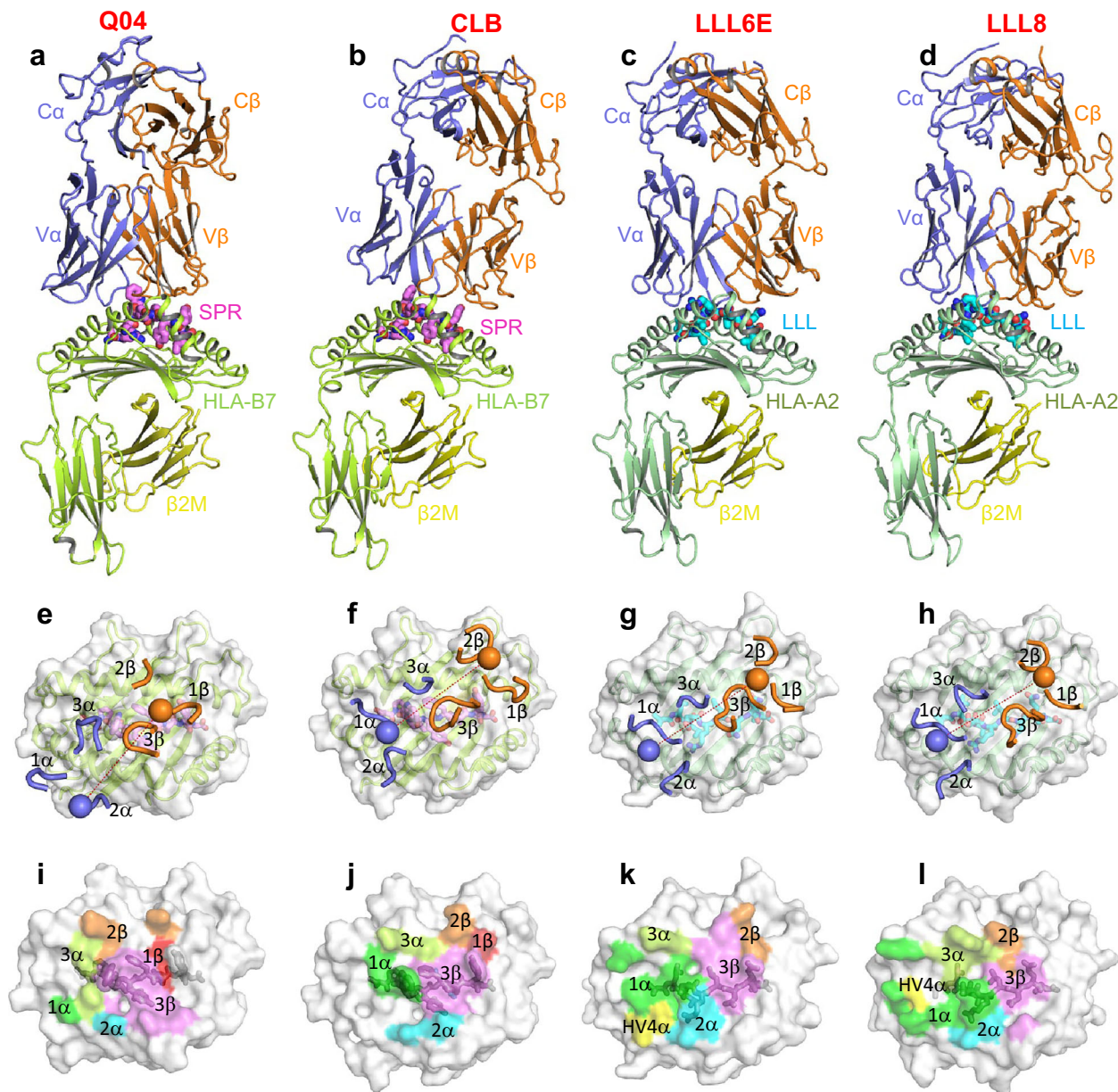


Fig. 2 | Structure of SPR-HLA-B7 and LLL-HLA-A2 in complex of TCRs. **a** Side view of Q04-SPR-HLA-B7 complex (ribbon diagram). TCR α chain, blue; TCR β chain, orange; HLA-B7 heavy chain, lime; β_2 -microglobulin (β_2m), yellow; SPR peptide, violet. **b** Side view of CLB-SPR-HLA-B7 complex. **c** Side view of LLL6E-LLL-HLA-A2 complex. TCR α chain, blue; TCR β chain, orange; HLA-A2 heavy chain, pale green; β_2m , yellow; LLL peptide, cyan. **d** Side view of LLL8-LLL-HLA-A2 complex (PDB accession code 8DNT)²⁵. **e** Positions of CDR loops of TCR Q04 on SPR-HLA-B7 (top view). CDRs of Q04 are shown as numbered blue (CDR1 α , CDR2 α , and CDR3 α) or orange (CDR1 β , CDR2 β , and CDR3 β) loops. HLA-B7 is depicted as a gray surface. The SPR peptide is drawn in violet in stick

representation. The blue and orange spheres mark the positions of the conserved intrachain disulfide of the V α and V β domains, respectively. The red dashed line indicates the crossing angle of TCR to pMHC. **f** Positions of CDR loops of TCR CLB on SPR-HLA-B7 (top view). **g** Positions of CDR loops of TCR LLL6E on LLL-HLA-A2 (top view). **h** Positions of CDR loops of TCR LLL8 on LLL-HLA-A2 (top view). **i** Footprint of TCR Q04 on SPR-HLA-B7. The top of the MHC molecule is depicted as a gray surface. The areas contacted by individual CDR loops are color-coded: CDR1 α , green; CDR2 α , cyan; CDR3 α , lime; HV4 α , yellow; CDR1 β , red; CDR2 β , orange; CDR3 β , violet. **j** Footprint of TCR CLB on SPR-HLA-B7 complex. **k** Footprint of TCR LLL6E on LLL-HLA-A2. **l** Footprint of TCR LLL8 on LLL-HLA-A2.

angles of TCR to pMHC⁴⁴ of 52° and 44°, respectively (Fig. 2e, f). The complexes also differ with respect to incident angle⁴⁵, which corresponds to the degree of tilt of TCR over pMHC: 20° for Q04 and 10° for CLB. In comparison with TCR-pMHC class I complexes from the Protein Data Bank (PDB) (130 complexes), the Q04 TCR complex has the 36th-highest crossing angle (72nd percentile), and the CLB TCR complex has the 63rd highest (52nd percentile). TCR LLL6E also docks symmetrically over LLL-HLA-A2 in a canonical diagonal orientation, with a crossing angle of 33° (Fig. 2g), which is nearly identical to that of TCR LLL8 (31°) (Fig. 2h), despite the use of different α/β chain pairs

(Supplementary Table 1). The incident angle of LLL6E is 10° compared to 3° for LLL8 (Fig. 2c, d).

As depicted by the footprints of Q04 and CLB on the pMHC surface (Fig. 2i, j), Q04 establishes contacts with the central portion of the SPR peptide mainly via the CDR3 α and CDR3 β loops, whereas CLB engages the central and C-terminal portions of SPR mostly through CDR1 α and CDR3 β . LLL6E contacts the N-terminal half of the LLL peptide primarily via CDR1 α and CDR2 α and the C-terminal half through CDR3 β (Fig. 2k). LLL8 makes a similar footprint on pMHC as LLL6E, except that CDR1 α and CDR3 α , rather than CDR1 α

and CDR2 α , mediate interactions with the N-terminal half of LLL (Fig. 2I).

Interaction of TCRs Q04 and CLB with HLA-B7

Of the total number of contacts (87) that TCR Q04 makes with HLA-B7, excluding the SPR peptide, CDR1 α , CDR2 α , and CDR3 α contribute 5%,

3%, and 52%, respectively, compared with 1%, 14%, and 25% for CDR1 β , CDR2 β , and CDR3 β , respectively (Fig. 3a, b and Table 1). Hence, CDR3 α accounts for considerably more of the binding interface with MHC than any other CDR. Residues Tyr93 α , Gly96 α , Thr97 α , and Tyr98 α of CDR3 α form a dense network of six hydrogen bonds with Arg62H and Gln65H of the HLA-B7 α 1 helix (Fig. 3a and Supplementary Table 3).

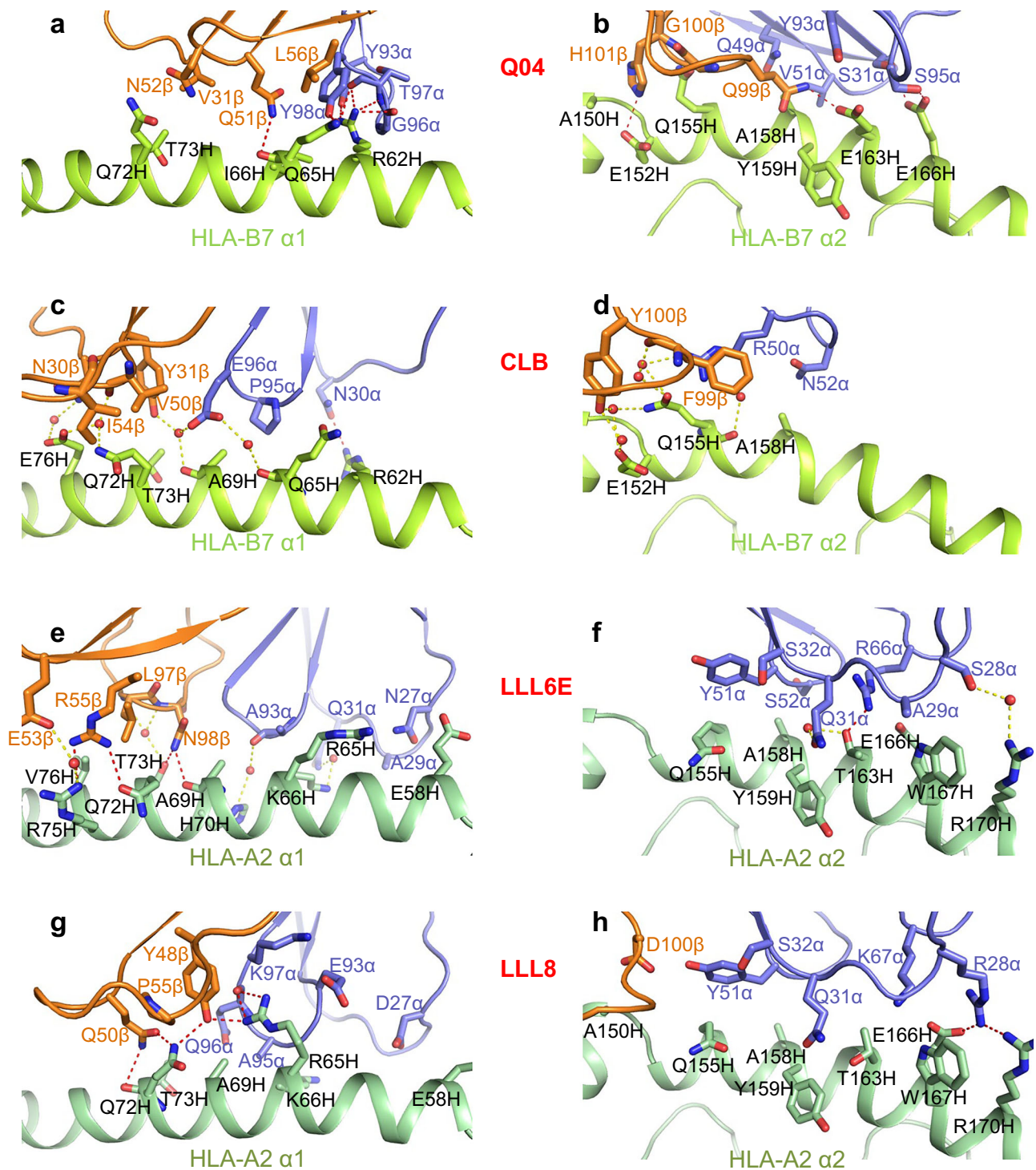


Fig. 3 | Interactions of TCRs with HLA-B7 and HLA-A2. a Interactions between TCR Q04 and the HLA-B7 α 1 helix. The side chains of contacting residues are drawn in stick representation with carbon atoms in blue (TCR α chain), orange (TCR β chain), lime (HLA-B7), or pale green (HLA-A2), nitrogen atoms in blue, and oxygen atoms in red. Hydrogen bonds are indicated by red dashed lines, water molecules are shown as red spheres, and water-mediated hydrogen bonds are indicated by

yellow dashed lines. **b** Interactions between Q04 and the HLA-B7 α 2 helix. **c** Interactions between CLB and the HLA-B7 α 1 helix. **d** Interactions between CLB and the HLA-B7 α 2 helix. **e** Interactions between LLL6E and the HLA-A2 α 1 helix. **f** Interactions between LLL6E and the HLA-A2 α 2 helix. **g** Interactions between LLL8 and the HLA-A2 α 1 helix. **h** Interactions between LLL8 and the HLA-A2 α 2 helix.

Table 1 | TCR CDR atomic contacts with peptide and MHC

TCRs	pMHC	α chain				β chain				Total ^a
		CDR1	CDR2	HV4	CDR3	CDR1	CDR2	HV4	CDR3	
# of contacts										
Q04	SPR	0	0	0	18	9	8	0	35	70
	HLA-B7	4	3	0	45	1	12	0	22	87
CLB	SPR	22	0	0	13	7	0	0	34	76
	HLA-B7	1	8	0	7	6	4	0	1	27
LLL6E	LLL	29	14	0	4	0	0	0	31	78
	HLA-A2	18	8	7	3	0	6	0	12	54
LLL8	LLL	25	5	0	16	1	2	0	15	64
	HLA-A2	19	5	4	30	0	30	0	4	92
% of contacts										
Q04	SPR	0	0	0	26	13	11	0	50	100
	HLA-B7	5	3	0	52	1	14	0	25	100
CLB	SPR	29	0	0	17	9	0	0	45	100
	HLA-B7	4	30	0	25	22	15	0	4	100
LLL6E	LLL	37	18	0	5	0	0	0	40	100
	HLA-A2	33	15	13	6	0	11	0	22	100
LLL8	LLL	39	8	0	25	2	3	0	23	100
	HLA-A2	21	5	4	33	0	33	0	4	100

Contacts were calculated between non-hydrogen atoms with a 4.0 Å distance cutoff.

^aTotal contacts reflect the total number of TCR–MHC or TCR–peptide contacts.

TCR Q04 interacts extensively with the HLA-B7 α 2 helix via CDR1 α , CDR2 α , CDR3 α , and CDR3 β (Fig. 3b). Overall, V α makes more contacts with MHC than V β (52 versus 35), as well as seven of 11 hydrogen bonds (Table 1 and Supplementary Table 3).

TCR CLB makes many fewer contacts with HLA-B7 than TCR Q04 (27 versus 87) (Table 1). Notably, the number of CLB TCR–MHC contacts is lower than all 130 Class I reference complexes from the PDB, for which the median number of TCR–MHC contacts is 79 and the lowest is 32 (PDB code 3UTS)⁴⁶. However, the relative paucity of direct CLB–HLA-B7 contacts may be compensated for, at least partially, by numerous water-mediated interactions, in particular seven water-mediated hydrogen bonds linking Arg50 α , Phe99 β , and Tyr100 β with Glu152H and Gln155H in the central section of the HLA-B7 α 2 helix (Fig. 3c, d and Supplementary Table 3). Six additional water-mediated hydrogen bonds link the SPR peptide to TCR CLB (Fig. 4f and Supplementary Table 5) (see below). We cannot say whether the Q04–SPR–HLA-B7 complex also contains water-mediated hydrogen bonds because the resolution of the Q04–SPR–HLA-B7 structure (2.75 Å) is insufficient to identify bound waters with confidence. Such identification requires a resolution of 2.5 Å or better, which is attained by the CLB–SPR–HLA-B7 complex (2.04 Å).

The contribution, if any, of bound waters to shape complementarity at the TCR–pMHC or other protein–protein interface may be quantified using the shape correlation statistic (S_c)⁴⁷, where $S_c = 1$ for interfaces with perfect geometric fit. The S_c value for the CLB–SPR–HLA-B7 complex is 0.83 with interfacial waters versus 0.75 without waters, indicating a substantial contribution to improving the overall fit. Thus, bound waters help correct imperfections in the CLB–SPR–HLA-B7 interface by filling cavities between TCR and pMHC, as well as by forming bridging hydrogen bonds to enhance polar interactions and neutralize unpaired hydrogen-bonding groups, as observed in other protein–protein complexes⁴⁸. By comparison, the S_c value for the Q04–SPR–HLA-B7 complex is 0.70 without waters, which is consistent with the similar K_{DS} of Q04 and CLB for SPR–HLA-B7 (0.43 μ M and 0.41 μ M, respectively).

Interaction of TCR LLL6E with HLA-A2

Of the total number of contacts (54) that TCR LLL6E makes with HLA-A2, excluding the LLL peptide, CDR1 α , CDR2 α , HV4 α , and CDR3 α contribute 33%, 15%, 13%, and 6%, respectively, compared with 0%, 11%, and 22% for CDR1 β , CDR2 β , and CDR3 β , respectively (Table 1). Hence, V α dominates the interactions of LLL6E with MHC (36 of 54 contacts; 67%), with the germline-encoded CDR1 α loop contributing more to MHC recognition than any other CDR (18 contacts). A similar degree of V α dominance (63% of contacts with MHC) is observed for TCR LLL8, but with the somatically generated CDR3 α loop making the greatest contribution (Table 1).

Although LLL6E and LLL8 use V α regions belonging to the same family (TRAV12-1 and TRAV12-2, respectively), the sequences of their germline-encoded CDR1 α and CDR2 α loops differ at several positions: ²⁷NSASQS³² and ²⁷DRGSQS³² (MHC-contacting residues underlined) for CDR1 α of LLL6E and LLL8, respectively, and ⁵⁰VYSSGN⁵⁵ and ⁵⁰IYSNGD⁵⁵ (MHC-contacting residues underlined) for CDR2 α of LLL6E and LLL8, respectively. Despite these differences, MHC-contacting residues that are conserved in TRAV12-1 and TRAV12-2 mediate similar interactions with MHC in the LLL6E–LLL–HLA-A2 and LLL8–LLL–HLA-A2 complexes (Fig. 3e–h and Supplementary Table 4). Thus, CDR1 α Gln31 contacts Tyr159H and Thr163H in both structures. Likewise, CDR2 α Tyr51 and CDR2 α Ser52 contact Gln155H and Ala158H in each complex. These conserved interactions serve as anchor points to enable TCRs LLL6E and LLL8 to dock onto LLL–HLA-A2 in similar orientations, as manifested by nearly identical crossing angles of 33° and 31°, respectively (Fig. 2g, h). This maintenance of germline-encoded interactions explains the interchangeability of TRAV12-1 and TRAV12-2 V α regions and supports the hypothesis of coevolution of TCR and MHC molecules^{49,50}. Superposition of the LLL8–LLL–HLA-A2 and LLL6E–LLL–HLA-A2 complexes gave an RMSD in α -carbon positions of 0.78 Å for the V α modules, showing that TRAV12-1 and TRAV12-2 dock very similarly on HLA-A2. Further supporting the interchangeability of TRAV12-1 and TRAV12-2, we used AlphaFold⁴³ to model the LLL6E–LLL–HLA-A2 complex with either TRAV12-1 or TRAV12-2 (see below). The two models were very similar and closely matched the LLL6E–LLL–HLA-A2 crystal structure.

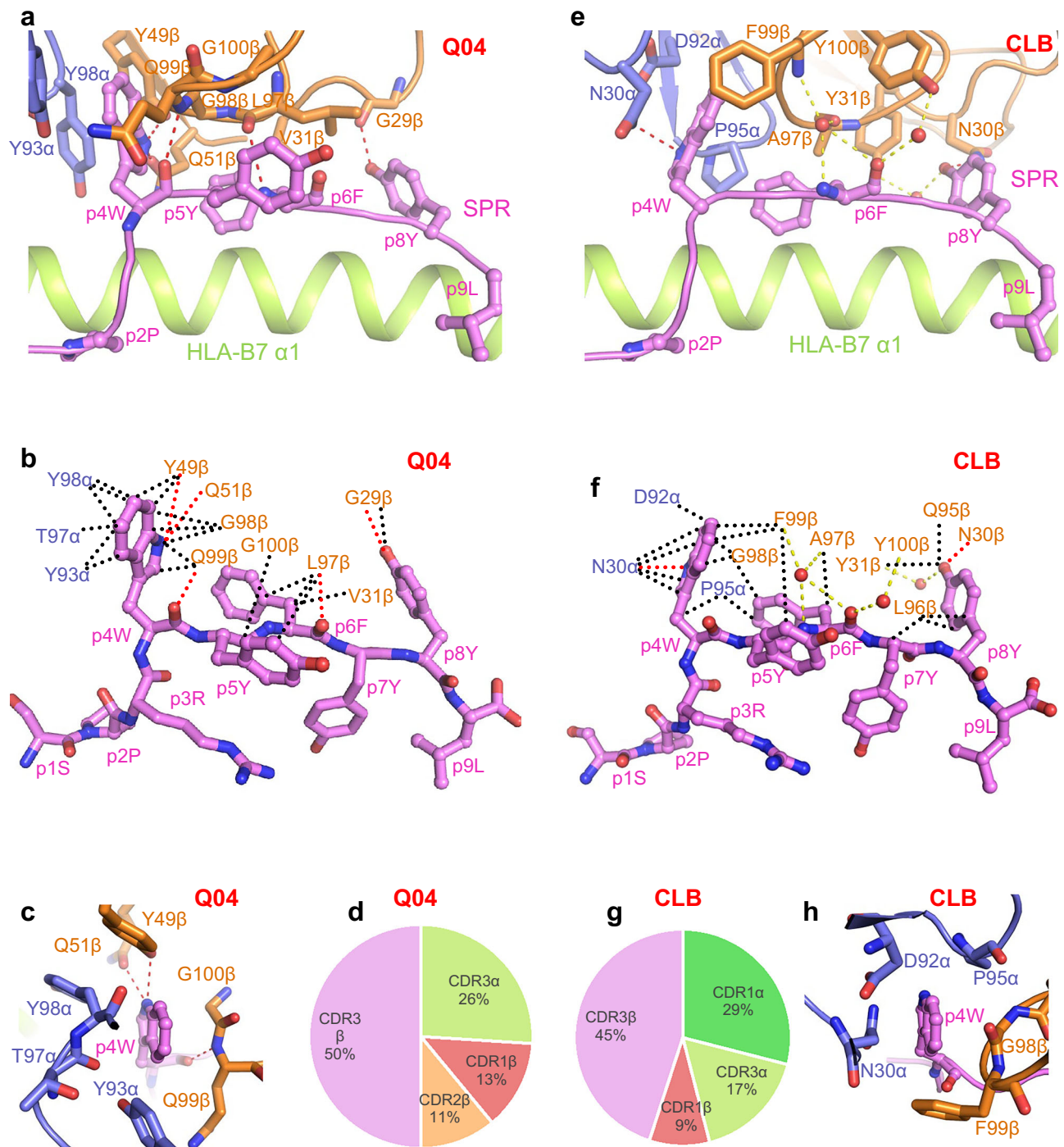


Fig. 4 | Interactions of SARS-CoV-2-specific TCRs with the SPR peptide.

a Interactions between TCR Q04 and the SPR peptide. The side chains of contacting residues are shown in stick representation with carbon atoms in blue (TCR α chain), orange (TCR β chain), or violet (SPR), nitrogen atoms in blue, oxygen atoms in red, and water molecules as red spheres. Peptide residues are identified by one-letter amino acid designation followed by position (p) number. Hydrogen bonds are indicated by red dashed lines. Water-mediated hydrogen bonds are drawn as yellow dashed lines. **b** Schematic representation of Q04–SPR interactions. Hydrogen bonds are red dotted lines, water-mediated hydrogen bonds are yellow dashed

lines, and van der Waals contacts are black dotted lines. For clarity, not all van der Waals contacts are shown. **c** Close-up of interactions of Q04 with P4 Trp of the SPR peptide. **d** Pie chart showing percentage distribution of TCR Q04 contacts to SPR peptide according to CDR in the Q04–SPR–HLA-B7 complex. **e** Interactions between TCR CLB and the SPR peptide. **f** Schematic representation of TCR CLB–SPR interactions. **g** Pie chart showing percentage distribution of TCR CLB contacts to SPR peptide according to CDR in the CLB–SPR–HLA-B7 complex. **h** Close-up of interactions of CLB with P4 Trp of the SPR peptide.

TCR LLL6E engages the HLA-A2 $\alpha 1$ helix through four hydrogen bonds linking Arg55 β and Asn98 β to Ala69H, Gln72H, Thr73H, and Arg75H (Fig. 3e and Supplementary Table 4). These interactions, which are not conserved in the LLL8–LLL–HLA-A2 complex due to utilization of an unrelated V β region (TRBV8 instead of TRBV7-2) (Fig. 3g), are

reinforced by a cluster of six water-mediated hydrogen bonds. The S_c value for the LLL6E–LLL6–HLA-A2 complex is 0.75 with interfacial waters versus 0.67 without waters ($\Delta S_c = 0.08$), while the S_c value for the CLB–SPR–HLA-B7 complex is 0.83 with interfacial waters versus 0.75 without waters ($\Delta S_c = 0.08$). Thus, interfacial waters make similar

positive contributions to improving shape complementarity in these two unrelated TCR–pMHC complexes.

The relatively low resolution of the LLL8–LLL–HLA-A2 structure (3.18 Å)²⁸ prevented us from identifying interfacial water molecules, and thus water-mediated hydrogen bonds, with a reasonable degree of accuracy. This limitation does not apply to the LLL6E–LLL–HLA-A2 structure, which we determined to considerably higher resolution (2.17 Å). Therefore, we cannot compare the number or nature of water-mediated hydrogen bonds in the two complexes.

SPR epitope recognition by TCRs Q04 and CLB

Upon binding SPR–HLA-B7, TCR Q04 buries 82% (357 Å²) of the peptide solvent-accessible surface. Q04 engages four residues in the central (P4 Trp, P5 Tyr, and P6 Phe) and C-terminal (P8 Tyr) portions of the SPR peptide, with a focus on P4 Trp (43 of 65 van der Waals contacts and three of five hydrogen bonds) (Supplementary Table 5 and Fig. 4a, b). Computational alanine scanning mutagenesis in Rosetta⁵¹ with the Q04–SPR–HLA-B7 structure indicates that P4 Trp indeed dominates the energetics of the interactions with TCR Q04 (Supplementary Table 6). In agreement with this prediction, we detected no interaction by surface plasmon resonance of Q04 with HLA-B7 presenting a mutant SPR peptide with alanine substitution at P4 Trp (Supplementary Fig. 2a). In the Q04–SPR–HLA-B7 structure, the side chain of P4 Trp inserts into a pocket formed by CDR3α, CDR2β, and CDR3β, where its indole ring forms two hydrogen bonds with the phenyl ring of CDR2β Tyr49 and the side chain of CDR2β Gln51 (Fig. 4c). Interactions between Q04 and the SPR peptide are mediated almost exclusively by CDR3α (26%) and CDR3β (50%) with minor contributions from CDR1β (13%) and CDR2β (11%) (Fig. 4d).

TCR CLB buries 84% (352 Å²) of the solvent-accessible surface of the SPR peptide upon binding SPR–HLA-B7. Of 76 total contacts that CLB establishes with SPR, CDR1α, CDR2α, and CDR3α account for 29%, 0%, and 17%, respectively, compared with 9%, 0%, and 45% for CDR1β, CDR2β, and CDR3β, respectively (Fig. 4e–g and Table 1). Similar to Q04, CLB engages five residues in the central (P4 Trp, P5 Tyr, and P6 Phe) and C-terminal (P7 Tyr and P8 Tyr) portions of the SPR peptide, with no contacts to the N-terminal portion (Supplementary Table 5 and Fig. 4e, f). However, the principal focus is on P4 Trp, which inserts into a pocket formed by Asn30α, Asp92α, Pro95α, Gly98β, and Phe99β, where it makes 43 van der Waals contacts, 20 of which involve Asn30α (Fig. 4h). Further strengthening the interaction between Asn30α and P4 Trp is a side-chain–side-chain hydrogen bond: CLB Asn30α Nδ2–Nε1 P4 Trp. Also important for recognition is P6 Phe, which engages TCR via five water-mediated hydrogen bonds: CLB Ala97β O–H₂O–O P6 Phe, CLB Ala97β O–H₂O–N P6 Phe, CLB Phe99β N–H₂O–O P6 Phe, CLB Ala99β N–H₂O–N P6 Phe, and CLB Tyr100β O_η–H₂O–O P6 Phe (Supplementary Table 5 and Fig. 4f). Computational alanine scanning⁵¹ with the CLB–SPR–HLA-B7 structure confirms that P4 Trp and P6 Phe dominate the binding energetics with TCR, along with P8 Tyr (Supplementary Table 6). In agreement with this prediction, we detected no binding of CLB to HLA-B7 loaded with a mutant SPR peptide with alanine at P4 Trp (Supplementary Fig. 2b). By comparison, the affinity (*K_D*) of CLB for HLA-B7 loaded with a mutant SPR peptide with alanine at P5 Tyr was 46 μM (Supplementary Fig. 2c), which is 112-fold lower than its affinity for wild-type SPR–HLA-B7 (0.41 μM). This affinity reduction is consistent with the unfavorable Δ*G* of 0.5 REU calculated using Rosetta⁵¹ (Supplementary Table 6), albeit somewhat higher in magnitude.

LLL epitope recognition by TCR LLL6E

Upon binding LLL–HLA-A2, LLL6E engages all eight solvent-exposed residues along the entire length of LLL, thereby burying 81% (321 Å²) of peptide solvent-accessible surface and enabling maximum readout of the peptide sequence. Of the 78 total contacts that LLL6E establishes with the LLL peptide, most (47; 60%) are mediated by Vα (Table 1). This

Vα bias, which is also a feature of the LLL8–LLL–HLA-A2 complex²⁸, allows pairing with multiple Vβs, which, like TRBV9 of LLL6E, are expected to make comparatively few interactions with the peptide, as well as MHC. CDR1α, CDR2α, and CDR3α account for 37%, 18%, and 5% of contacts, respectively, compared to 0%, 0%, and 40% for CDR1β, CDR2β, and CDR3β, respectively (Fig. 5a–c). Although Vα dominates the interactions of both LLL6E and LLL8 with both MHC and peptide (~65% of total contacts), Vβ also makes significant contributions to recognition. Most notably, CDR3β of LLL6E alone accounts for 40% of contacts with LLL (Fig. 5c), which is more than any other CDR, while CDR3β of LLL8 accounts for 23% of such contacts (Fig. 5h).

As noted above, the germline-encoded CDR1α and CDR2α loops of LLL6E and LLL8, which use TRAV12-1 and TRAV12-2 regions, respectively, differ significantly in sequence: NSASQS and DRGSQS (peptide-contacting residues underlined) for CDR1α of LLL6E and LLL8, respectively, and VYSSGN and IYSNGD (peptide-contacting residues underlined) for CDR2α of LLL6E and LLL8, respectively. Nevertheless, peptide-contacting residues that are conserved in TRAV12-1 and TRAV12-2 (CDR1α Gln31, CDR1α Ser32, and CDR2α Tyr51) mediate very similar interactions with the LLL peptide, as well as with MHC (see above), in the LLL6E–LLL–HLA-A2 and LLL8–LLL–HLA-A2 complexes (Fig. 5b, g and Supplementary Table 7). In particular, CDR1α Gln31 and CDR1α Ser32 engage P2 Leu, P4 Asp, and P5 Arg in the N-terminal half of the LLL epitope in both structures. In the LLL6E–LLL–HLA-A2 complex, these CDR1α residues form a cluster of six hydrogen bonds with peptide that are mostly conserved (4 of 6) in the LLL8–LLL–HLA-A2 complex: Gln31α Nε2–O P2 Leu, Gln31α Nε2–N P4 Asp, Gln31α O–N_η1 P5 Arg, Ser32α O_γ–Oδ1 P4 Asp, Ser32α O_γ–Oδ2 P4 Asp, and Ser32α O_γ–N_η1 P5 Arg (Fig. 5e). These direct interactions are reinforced by four water-mediated hydrogen bonds (Supplementary Table 7). Nine additional water bridges link CDR3β to P6 Leu and P8 Gln. The low resolution of the LLL8–LLL–HLA-A2 complex (3.18 Å) precluded the identification of interfacial water molecules for comparison²⁸.

LLL-specific TCRs from COVID-19 convalescent patients are characterized by dominant usage of TRAV12-1 and TRAV12-2 gene segments (>50%) with no observed usage of TRAV12-3²⁷. TRAV12-1 and TRAV12-2, which we have shown are functionally interchangeable, both encode CDR1α residues Gln31α and Ser32α, whereas TRAV12-3 encodes CDR1α residues Gln31α and Tyr32α. Computational mutagenesis using Rosetta⁵¹ of Ser32α to Tyr in the LLL6E–LLL–HLA-A2 and LLL8–LLL–HLA-A2 complexes gave highly unfavorable Δ*G* values of 5.2 REU and 3.4 REU, respectively, indicating that TRAV12-3-encoded CDR1α Tyr32 would be incompatible with the LLL6E/LLL8 mode of LLL–HLA-A2 engagement. In agreement with prediction, we detected no binding by surface plasmon resonance of TCR LLL6E with the Ser32α to Tyr mutation to LLL–HLA-A2 (Supplementary Fig. 2d).

The CDR3α of LLL6E (and LLL6) differs in sequence from LLL8, and does not possess the previously noted LLL TCR subsequence motif (G/N)(G/A)(Q/N)K exemplified by LLL8 (GAQK)²⁸, although its equivalent subsequence (AGNK) nearly matches the reported motif. Structurally, the LLL6E CDR3α loop engages the pMHC in a similar mode as LLL8 CDR3α loop, and most pMHC contacts are from small amino acids at the loop apex (Ala93α, Gly94α) (Fig. 5a), as with LLL8 (Gly94α, Ala95α in that case).

In contrast to CDR3α of LLL8, which accounts for 25% of contacts with the LLL peptide (Fig. 5h), CDR3α of LLL6E contributes only 5% of contacts (Fig. 5c). Conversely, CDR3β of LLL6E mediates 40% of contacts with LLL compared to only 23% by CDR3β of LLL8. These interactions include three direct and nine water-mediated hydrogen bonds with P6 Leu and P8 Gln that anchor TCR LLL6E to the C-terminal half of the LLL peptide (Fig. 5d and Supplementary Table 7). While the CDR3β loop of LLL6E shares no obvious sequence features with LLL8, both interfaces feature a negatively charged residue (Glu96β and Asp97β for LLL6E and LLL8, respectively) engaging the LLL peptide backbone

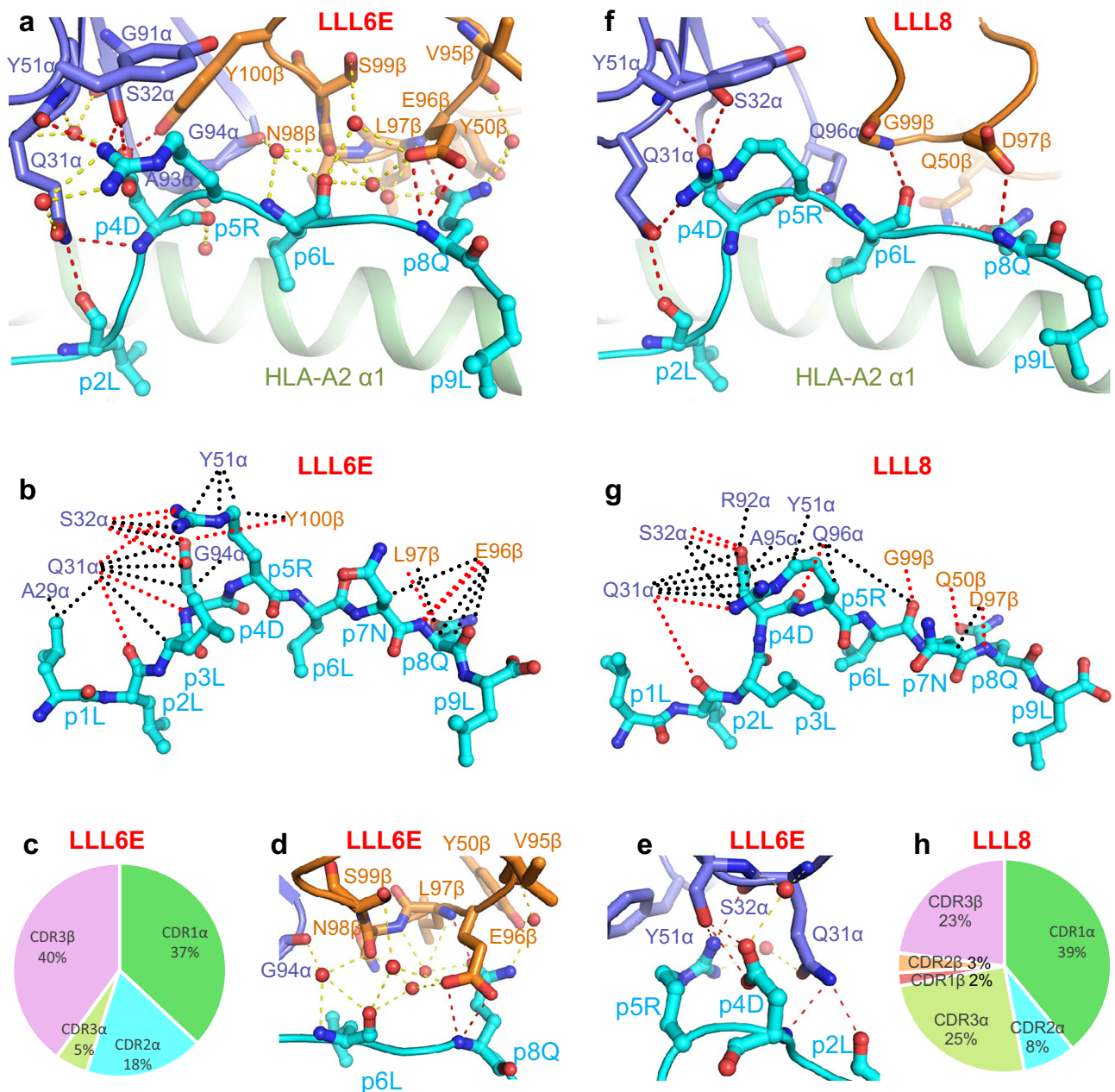


Fig. 5 | Interactions of SARS-CoV-2-specific TCRs with the LLL peptide.

a Interactions between TCR LLL6E and the LLL peptide. The side chains of contacting residues are shown in stick representation with carbon atoms in blue (TCR α chain), orange (TCR β chain), or cyan (LLL), nitrogen atoms in blue, oxygen atoms in red, and water molecules as red spheres. **b** Schematic representation of LLL6E–LLL interactions. Hydrogen bonds are red dotted lines, water-mediated hydrogen bonds are yellow dashed lines, and van der Waals contacts are black dotted lines. For clarity, not all van der Waals contacts are shown. **c** Pie chart showing percentage

distribution of TCR LLL6E contacts to LLL peptide according to CDR in the LLL6E–LLL–HLA-A2 complex. **d** Close-up of water bridges linking LLL6E to P6 Leu and P8 Gln of the LLL peptide. **e** Close-up of interactions of LLL6E with P4 Asp and P5 Arg of the LLL peptide. **f** Interactions between TCR LLL8 and the LLL peptide. **g** Schematic representation of LLL8–LLL interactions. **h** Pie chart showing percentage distribution of TCR LLL8 contacts to LLL peptide according to CDR in the LLL8–LLL–HLA-A2 complex.

at the same site (Fig. 5a, f), representing predicted energetic hotspots (Supplementary Table 8)²⁸.

Cross-recognition of SPR and LLL variants and homologous epitopes

The structures of TCRs bound to SPR–HLA-B7 and LLL–HLA-A2 provide frameworks for understanding T cell recognition of viral variants and homologous epitopes from other human coronaviruses. We assembled a set of SPR and LLL nucleocapsid epitope sequences from five representative coronaviruses: SARS-CoV-1, OC43, HKU1, NL63, and

229E (Supplementary Table 9). Computational mutagenesis in Rosetta⁵¹ was used to predict the effects on binding of TCRs Q04, CLB, and LLL6E ($\Delta\Delta G$). This modeling protocol was previously found to be accurate in estimating $\Delta\Delta G$ s in other TCR–pMHC complexes^{42,52,53}.

The predicted ability of SARS-CoV-2-specific TCRs Q04 and CLB to recognize peptides from other coronaviruses homologous to the SPR epitope varied considerably. Whereas SARS-CoV-2 and SARS-CoV-1 share an identical SPR epitope (SPRWYFYLYL), the homologous peptides from OC43 and HKU1 betacoronaviruses (LPRWYFYLYL; amino acid replacement in bold) differ at P1 with a serine-to-leucine

substitution. However, P1 Ser does not contact Q04 or CLB in the crystal structures (Fig. 4b, f), nor is it an anchor residue for HLA-B7. No significant effect of this substitution on TCR binding was predicted by computational mutagenesis⁵¹ ($\Delta\Delta G$ values of -0.1 and 0.2 Rosetta Energy Units (REU) for Q04 and CLB, respectively) (Supplementary Table 9), implying possible cross-recognition of OC43 and HKU1 by these TCRs. To validate this prediction, we measured the binding of Q04 and CLB to HLA-B7 loaded with the LPRWFYYL peptide. Q04 bound LPRWFYYL-HLA-B7 with $K_D = 0.12 \mu\text{M}$, which is actually -threefold higher affinity than for SPRWFYYL-HLA-B7 ($K_D = 0.43 \mu\text{M}$) (Supplementary Fig. 2e). CLB bound LPRWFYYL-HLA-B7 with $K_D = 0.09 \mu\text{M}$, which is -fourfold higher affinity than for SPRWFYYL-HLA-B7 ($K_D = 0.41 \mu\text{M}$) (Supplementary Fig. 2f). Thus, the Ser to Leu substitution at P1 did not diminish TCR binding.

The homologous epitopes of NL63 and 229E alphacoronaviruses (PPKVVHFYYL and SPKLHFYYL, respectively) differ at residues P4 and P5, which contact TCR in the Q04-SPR-HLA-B7 and CLB-SPR-HLA-B7 complexes (Fig. 4b, f). Large disruptive effects on TCR affinity ($\Delta\Delta G > 3.3$ REU) were predicted for both peptides (Supplementary Table 9), suggesting no cross-recognition of NL63 or 229E by Q04 or CLB. These results are consistent with functional assays showing that SARS-CoV-2-specific CD8⁺ T cells can be activated by antigen-presenting cells pulsed with OC43 or HKU1 homologous peptides, but not NL63 or 229E peptides²⁹.

SARS-CoV-2 and SARS-CoV-1 share an identical LLL epitope (LLLDRLNQL) (Supplementary Table 9). However, homologous peptides from OC43 (LVLAKLGKD), HKU1 (LVLAKLGKD), NL63 (AVNLALKNL), and 229E (AVNLALKSL) differ at 6 or 7 positions, most notably P4 and P5, which form extensive contacts with SARS-CoV-2-specific TCRs LLL6E and LLL8 in the complex structures (Fig. 5b, g). Computational mutagenesis⁵¹ predicted disruption of TCR binding ($\Delta\Delta G > 4.0$ REU) for all four peptides, making cross-recognition of OC43, HKU1, NL63, or 229E by LLL6E or LLL8 unlikely.

Based on SARS-CoV-2 sequences in the GISAID database (<https://www.gisaid.org>)⁵⁴, both the SPR and LLL epitopes are highly conserved (Supplementary Table 10). The SPR polymorphism with the highest frequency (0.004%) is L113I at P9, a primary MHC anchor position at the peptide C-terminus that does not contact TCR Q04 or CLB. Since the conservative leucine-to-isoleucine replacement is also not likely to affect SPR binding to HLA-B7, the L113I polymorphism should not impact TCR recognition. To test this prediction, we measured the binding of TCRs Q04 and CLB to HLA-B7 loaded with SPR peptide bearing the L113I polymorphism. Both Q04 and CLB bound L113I-HLA-B7 with $K_D = 0.11 \mu\text{M}$ (Supplementary Fig. 2g, h), which is -fourfold higher affinity than for SPR-HLA-B7 ($K_D = 0.43 \mu\text{M}$ for Q04 and $0.41 \mu\text{M}$ for CLB). Thus, the Leu to Ile substitution at P9 did not reduce TCR recognition. Similar considerations apply to other SPR epitope variants (Supplementary Table 10). Moreover, the low frequency of SPR variants makes them unlikely to be encountered by the population, including HLA-B7 individuals, either following vaccination or infection. By contrast, LLL epitope variants occur at much higher frequency, up to 3.34% for the Q229K polymorphism at TCR-contacting position P8 (present in 564,921 out of 16,931,861 nucleocapsid sequences), and 0.12% for the L230F polymorphism at primary MHC anchor position P9 (Supplementary Table 10). The LLL mutant Q229K was recently reported to be present in BA.2.86/JN.1 SARS-CoV-2 variants as a T cell escape hotspot⁵⁵, which is in accordance with its relatively high prevalence in GISAID sequences, and the Q229K substitution is predicted to be disruptive for LLL6E TCR binding in Rosetta ($\Delta\Delta G = 1.3$ REU; Supplementary Table 9). In agreement with prediction, we detected no binding by surface plasmon resonance of TCR LLL6E to HLA-A2 loaded with the Q229K peptide (Supplementary Fig. 2i). Accordingly, superposition of the LLL^{Q229K}-HLA-A2 structure⁵⁵ onto the LLL6E-LLL-HLA-A2 and LLL8-LLL-HLA-A2 complexes revealed that the Q229K substitution would result in the loss of a main-

chain-side-chain hydrogen bond with LLL6E (Leu97 β N-O ϵ 1 P8 Gln) and a side-chain-side-chain hydrogen bond with LLL8 (Gln50 β N ϵ 2-O ϵ 1 P8 Gln), as well as the introduction of a positively charged residue in the interface (Supplementary Fig. 3). In addition, previous modeling of the L230F substitution indicated that it would likely prevent epitope presentation by HLA-A2²⁸.

Conformational changes in pMHC upon TCR binding

To identify possible conformational changes in SPR-HLA-B7 induced by TCR binding, we determined the structure of unbound SPR-HLA-B7 to 1.98 Å resolution (Supplementary Table 2). Unambiguous electron density was observed for the MHC-bound peptide (Supplementary Fig. 4). We first compared our SPR-HLA-B7 structure with one reported previously for the same ligand but in a different space group²⁹. Although the two structures are nearly identical (RMSD of 0.64 Å for main-chain atoms of the MHC α 1/ α 2 module and SPR peptide), the side chain of P5 Tyr adopts different conformations characterized by a 120° flip about the C α -C β axis (Fig. 6a). The different conformations of the P5 Tyr side chain in the two unbound SPR-HLA-B7 structures do not appear to be due to differences in crystal packing because P5 Tyr does not contact neighboring molecules in either crystal lattice. Superposition of the MHC α 1 α 2 domains of the unbound SPR-HLA-B7 structures onto those of SPR-HLA-B7 in complex with TCR Q04 or CLB showed that the conformation of P5 Tyr in the Q04-SPR-HLA-B7 or CLB-SPR-HLA-B7 complex is similar to its conformation in the previously reported unbound SPR-HLA-B7 structure²⁹ (Fig. 6b, c), but different from its conformation in the unbound SPR-HLA-B7 structure reported here, implying that the TCRs are selecting an alternative conformation of P5 Tyr for docking rather than inducing a conformational change. By contrast, the side chain of P4 Trp, which adopts similar conformations in the two unbound SPR-HLA-B7 structures (Fig. 6a), rotates 180° about the C α -C β axis to accommodate Tyr49 β and Gln51 β of Q04 (Fig. 6b) and 150° about the C α -C β axis to accommodate Phe99 β of CLB (Fig. 6c), indicating TCR-induced conformational changes in both cases.

Superposition of the MHC α 1 α 2 domains of unbound LLL-HLA-A2 (7KGQ)⁵⁶ onto those of LLL-HLA-A2 in complex with TCR LLL6E showed small yet relevant differences in peptide conformation, corresponding to an RMSD of 0.6 Å for main-chain atoms of LLL. The largest displacement by far is for P5 Arg, whose α -carbon shifted 2.4 Å and whose side chain moved 7.2 Å to allow hydrogen bond formation with Gln31 α and Ser32 α of CDR1 α (Fig. 6d).

Modeling Q04, CLB, and LLL6E TCR-pMHC complexes with AlphaFold

While AlphaFold has been able to model immune recognition by antibodies and TCRs with high accuracy in some cases, it has demonstrated limited overall success for those complexes⁵⁷⁻⁵⁹. To investigate modeling performance for previously unseen TCR-pMHC complexes, we used AlphaFold2 (AlphaFold-Multimer v.2.3^{60,61}, in the TCRmodel2 protocol⁵⁹) and AlphaFold3⁶² to model the three TCR-pMHC complex structures determined in this study. For each complex, 1000 predictions were generated from sequence, and the top-ranked model from each method was assessed for accuracy. To prevent differences in available structural templates, both methods were run with a PDB template date cutoff of September 30, 2021.

Modeling accuracy assessments indicate variable performance for the different complexes (Table 2). The Q04-SPR-HLA-B7 complex was modeled with high accuracy using both TCRmodel2 (AlphaFold2) and AlphaFold3, with sub-Ångstrom interface residue RMSD between models and corresponding X-ray structure. In contrast, neither modeling method generated highly accurate top-ranked models for the CLB-SPR-HLA-B7 complex, which contained relatively high model confidence and/or interface pLDDT (l-pLDDT) scores (noted in bold in Table 2). The LLL6E-LLL-HLA-A2 complex,

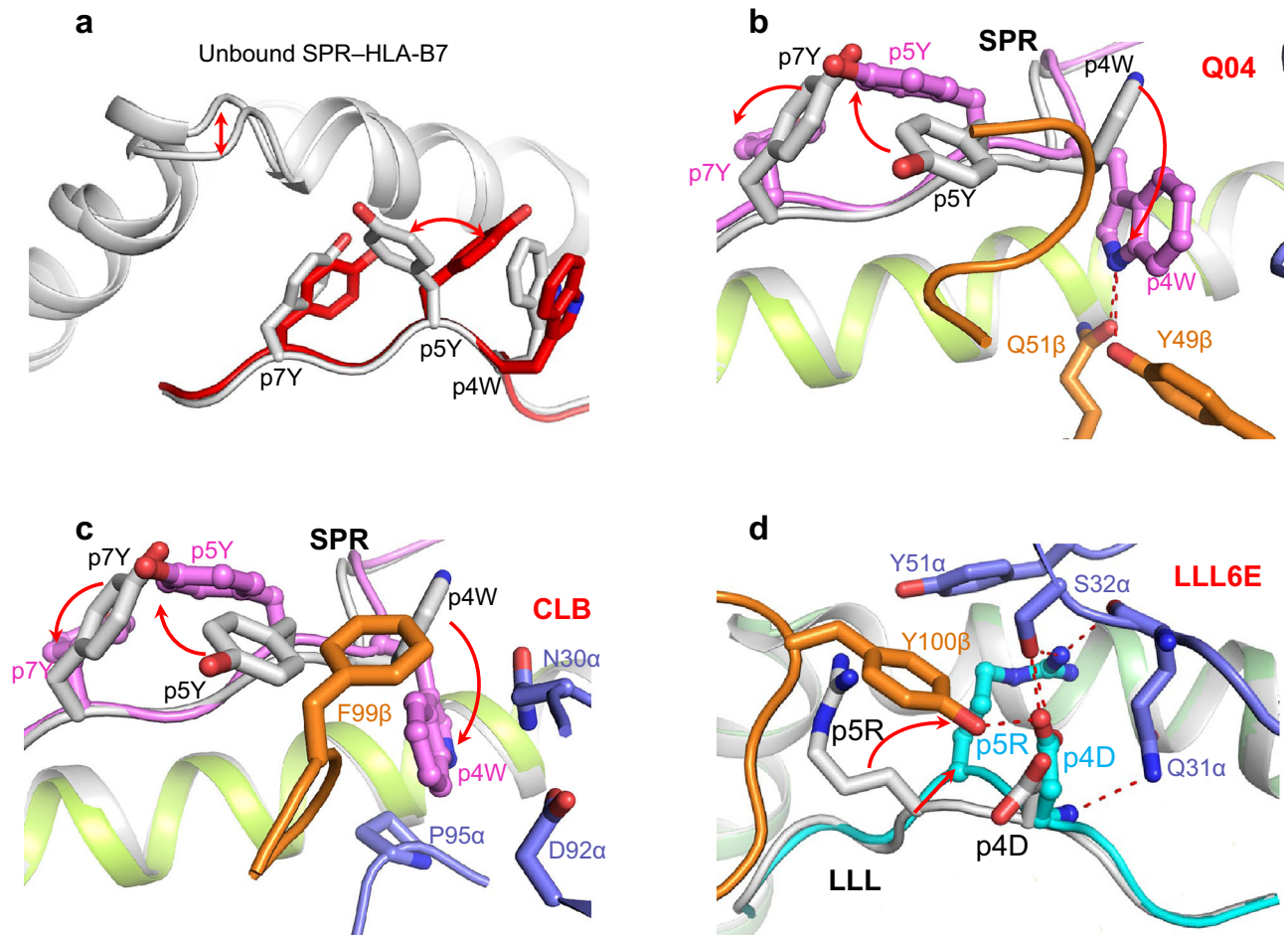


Fig. 6 | Conformational changes in pMHC upon TCRs binding. **a** Side view of unbound SPR-HLA-B7 in this study superposed on unbound SPR-HLA-B7 from a previous study (7LGD)²⁹. Double-headed red arrows indicate sites of structural difference. Carbon atoms of the superposed SPR peptides are gray and red (this study); nitrogen atoms are blue; oxygen atoms are red. HLA-B7 is gray. Residue labels for SPR are aligned with the α -carbon atom of the respective residue. **b** Superposition of SPR-HLA-B7 in unbound form (7LGD)²⁹ and bound to TCR Q04 showing rearrangements of P4 Trp, P5 Tyr, and P7 Tyr of the SPR peptide in the Q04-SPR-HLA-B7 complex. (unbound SPR-HLA-B7, gray; bound SPR, violet;

bound HLA-B7, lime; TCR α chain, blue; TCR β chain, orange). Single-headed red arrows indicates structural shift induced by TCR binding. **c** Superposition of SPR-HLA-B7 in unbound form (7LGD)²⁹ and bound to TCR CLB showing rearrangements of P4 Trp, P5 Tyr, and P7 Tyr of SPR in the CLB-SPR-HLA-B7 complex. **d** Superposition of LLL-HLA-A2 in unbound form (7KGQ)⁵⁶ and in complex with TCR LLL6E showing displacement of P5 Arg of the LLL peptide induced by LLL6E binding (unbound LLL-HLA-A2, gray; bound LLL, cyan; bound HLA-A2, green; TCR chain α , blue; TCR chain β , orange).

which was modeled separately with either V α TRAV12-1 or TRAV12-2 for LLL6, was modeled accurately by both TCRmodel2 and AlphaFold3, with medium and high accuracy models from TCRmodel2 and AlphaFold3, respectively, for the LLL6 complexes. As with the Q04-SPR-HLA-B7 models, the LLL6 and LLL6E complex models with high CAPRI accuracy had <1 Å interface RMSD from the X-ray structure, and relatively high AlphaFold confidence based on model confidence and I-pLDDT values. The range of model accuracies is evident when comparing representative modeled structures with the X-ray structures (Fig. 7), which shows large divergence in binding mode for the CLB TCR with respect to the experimentally determined structure, and recapitulation of the binding mode for the Q04 and LLL6E complexes.

Comparison of the modeled LLL6 interface from AlphaFold3 with the X-ray complex structure (Fig. 8) shows generally accurate backbone and side chain conformations in the model, with the exception of the peptide Arg5 side chain. Also shown in Fig. 8 is the modeled complex interface with the native TRAV12-2 LLL6 TCR, which indicates the positions and predicted side chain conformations of three interface-proximal CDR1 α and CDR2 α residues that differ from those in the TRAV12-1 LLL6 complex X-ray structure.

Due to the observed variability in predictive accuracy among the modeled TCR-pMHC interfaces, we investigated additional determinants of AlphaFold modeling success. As success focused on the top-ranked model in Table 2, we explored whether lower-ranked models from TCRmodel2 or AlphaFold3 contained more accurate models, particularly for the CLB complex (Supplementary Table 11). Model accuracy levels increased from medium to high for TCRmodel2 when considering the full set of 1000 models for LLL6E and LLL6, while for the CLB complex, both algorithms generated medium accuracy models within their sets of 1000, reflecting higher accuracy than the top-ranked models (incorrect and acceptable accuracy for TCRmodel2 and AlphaFold3, respectively). This highlights some room for improvement in scoring and model ranking, observed before in TCR-pMHC and antibody-antigen benchmarking using TCRmodel2 and AlphaFold2^{58,59}, where in principle, top-ranked model accuracy would be improved by identification of accurate versus inaccurate modeled complexes.

In addition to model ranking, we also explored the accuracy of these AlphaFold-based methods for modeling TCRs in complex with other SARS-CoV-2 epitopes. Three TCR-pMHC complex structures, all containing different epitopes from the spike glycoprotein, were

Table 2 | AlphaFold3 and TCRmodel2 TCR–pMHC complex modeling accuracy

TCR complex ^a	Modeling protocol ^b	Model confidence ^c	I-pLDDT ^c	I-RMSD, Å ^d	L-RMSD, Å ^d	DockQ ^d	Accuracy ^d
Q04	TCRmodel2	0.903	90.0	0.86	1.31	0.869	High
	AF3	0.922	94.2	0.68	1.13	0.902	High
CLB	TCRmodel2	0.909	93.1	7.22	13.84	0.105	Incorrect
	AF3	0.846	88.8	2.59	6.32	0.481	Acceptable
LLL6E (TRAV12-1*02)	TCRmodel2	0.873	86.8	1.51	3.2	0.696	Medium
	AF3	0.902	92.8	0.98	2.88	0.777	High
LLL6 (TRAV12-2*01)	TCRmodel2	0.879	89.5	1.09	2.89	0.773	Medium
	AF3	0.912	92.5	0.71	1.27	0.89	High

^aTCR–pMHC complex structure modeled, denoted by TCR name: Q04–SPR–HLA-B7 (Q04), CLB–SPR–HLA-B7 (CLB), and LLL6–LLL–HLA-A2 (LLL6). For LLL6, the complexes were separately modeled using TRAV12-1*02 (“LLL6E”) or TRAV12-2*01 (“LLL6”) for the TCR α chain, as indicated.

^bModeling protocol used, either TCRmodel2 or AlphaFold3 (AF3). For each protocol, 1000 models were generated per complex, with a template date cutoff of 9/30/2021, and the scores and accuracies shown are for the top-ranked model (based on model confidence score) for the given complex and protocol.

^cAlphaFold model confidence score ($0.8 \times \text{ipTM} + 0.2 \times \text{pTM}$) and interface pLDDT (I-pLDDT) score for the top-ranked model. Values in bold denote relatively high scores associated with accurate models based on previous analysis (model confidence > 0.875 or I-pLDDT > 87.5)⁵³.

^dInterface RMSD (I-RMSD), ligand RMSD (L-RMSD), DockQ score, and accuracy (based on CAPRI criteria) were calculated using the DockQ program⁸⁹ based on comparison of the top-ranked model with the corresponding X-ray structure. I-RMSD and L-RMSD were calculated using backbone atoms.

identified in the PDB with release dates after the September 2021 AlphaFold2 and AlphaFold3 training set cutoff. Of note, one of the complexes has related complex structures containing the same epitope (YLQ) that were released in the PDB before that date cutoff^{39,40}, thus the related complexes could have been seen by AlphaFold2 or AlphaFold3 during training. These three complexes were modeled using TCRmodel2 and AlphaFold3 using the same protocols used for the SPR and LLL epitope-containing complexes, and top-ranked models from both methods were assessed for accuracy (Supplementary Table 12). Accuracies varied slightly across the complexes and algorithms, with one of the three complexes achieving a high accuracy model (PDB code 8RJ5, with TCRmodel2), while the other two complexes achieved medium accuracy models with both methods. While the size of this additional set (three complexes) does not permit comparisons of AlphaFold modeling accuracy for different epitopes or source proteins, our results support previous benchmarking highlighting strong but variable accuracy AlphaFold2 and AlphaFold3 modeling accuracy for TCR–pMHC complexes^{42,59}.

Overall, the modeling of these previously unseen complexes underscores the capability of both AlphaFold-based methods to generate accurate TCR–pMHC models in some cases, along with the limitations in accuracy and scoring that indicate the need for further developments to consistently generate high quality models.

Discussion

Most SARS-CoV-2 N protein T cell epitopes are located in the RNA-binding (e.g., SPR) or dimerization domains of this structural protein⁶³. LLL, by contrast, is located in the central Ser/Arg-rich linker region connecting these domains that regulates their RNA-binding and dimerization activities. Because N protein epitopes, unlike S protein epitopes, are highly conserved, N-specific T cells show equivalent cross-reactivity against ancestral SARS-CoV-2 and VOCs^{23,64–70}. Thus, N-specific T cells may constitute a critical second line of defense following the antibody response for providing long-term protection against SARS-CoV-2 variants.

A striking feature of the T cell response to the HLA-B*07:02-restricted SPR epitope is its clonal diversity^{29–31}, whereby TCRs employ promiscuous α/β chain pairing to bind SPR–HLA-B7 in structurally different ways. This is illustrated by the Q04–SPR–HLA-B7 and CLB–SPR–HLA-B7 complexes reported here in which Q04 and CLB dock onto SPR–HLA-B7 with different crossing angles (52° and 44°, respectively) and different incident angles (20° and 10°, respectively). The clonal, and therefore structural, diversity of SPR-specific TCRs may enable them to circumvent epitope mutations that might otherwise

disrupt TCR recognition. Indeed, the frequencies of SPR epitope variants in the GISAID database⁵⁴ are exceedingly low (< 0.005% per variant) and are lower than previously described frequencies for spike T cell epitope substitutions (e.g., P272L:0.56%; T1006I:0.04%)⁴⁰. This apparent lack of dissemination of SPR variants in the wild suggests that they do not confer a selective advantage to the virus and/or that the diversity of SPR-specific TCRs prevents variant spread. Similar to SPR-specific TCRs, TCRs specific for the HLA-A*02:01-restricted spike epitope RLQ are highly diverse⁵⁵. We found that, whereas some RLQ-specific TCRs were unable to tolerate the most common natural mutation in RLQ (T1006I), recognition by other RLQ-specific TCRs was unaffected^{40,42}. Structural analysis of TCR–RLQ–HLA-A2 complexes showed that there are multiple solutions to recognizing RLQ, as there are for recognizing SPR, and that collectively these solutions can probably circumvent a wide variety of epitope mutations. In this way, CD8⁺ with diverse TCR repertoires can generate broadly protective immune responses that are often capable of recognizing both the wild-type virus and newly emerging variants^{71–74}.

TCRs specific for the HLA-A*02:01-restricted LLL epitope²⁸ are much less structurally diverse than HLA-B*07:02-restricted, SPR-specific TCRs^{29–31}, particularly with respect to the α chain. The majority of LLL-specific TCRs use the almost identical TRAV12-1 or TRAV12-2 gene segments, which dominate interactions with MHC²⁸. Conserved interactions between the germline-encoded CDR1 α and CDR2 α loops of these two V α regions and MHC explain the nearly identical docking topologies of the LLL6E–LLL–HLA-A2 and LLL8–LLL–HLA-A2 complexes. Opposite to SPR-specific TCRs, the restricted structural diversity of LLL-specific TCRs may facilitate viral escape from T cells targeting this epitope. In this regard, the frequency of LLL epitope variants in the GISAID database⁵⁴ is much higher than that of SPR variants. The mutation with the highest frequency (3.34%) is Q229K at TCR-contacting position P8. This mutation does not affect peptide binding to HLA-A2⁵⁵, but is predicted to disrupt binding by both LLL6E and LLL8. Consistent with this prediction, the Q229K mutation in Omicron variant BA.2.86/JN.1 was very recently found to evade T cell immunity⁵⁵. The high frequency of the Q229K mutation may therefore reflect dissemination in the wild due to escape from T cell surveillance.

Although the low frequency of SPR epitope variants may be attributable to T cell control, it is also possible that the intrinsic conserved components of the viral protein, such as highly networked epitopes in HIV structural proteins⁷⁵, could play a role. Conversely, the much higher frequency of LLL epitope variants may be due not only to lack of T cell control but also to the intrinsic plasticity of the viral protein.

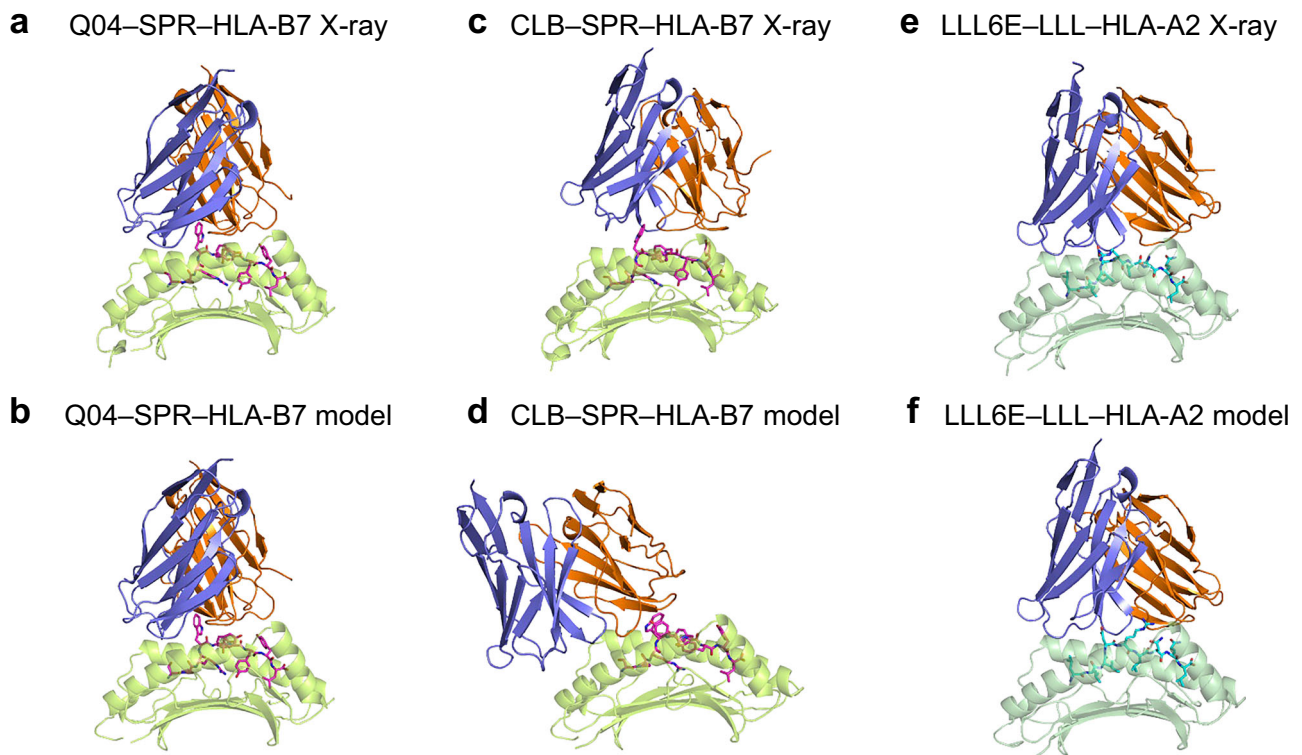


Fig. 7 | Structural models of Q04, CLB, and LLL6E TCR-pMHC complexes in comparison with X-ray structures. X-ray and modeled structures of Q04-SPR-HLA-B7 (**a, b**), CLB-SPR-HLA-B7 (**c, d**), LLL6E-LLL-HLA-A2 (**e, f**) are shown, with models from AlphaFold3 (Q04-SPR-HLA-B7, LLL6E-LLL-HLA-A2) or

TCRmodel2 (CLB-SPR-HLA-B7) and represent top-scoring models based on model confidence score. TCRs are colored blue (α chain) and orange (β chain), peptides are magenta (SPR peptide) or cyan (LLL peptide), and MHCs are colored light green (HLA-B7) or dark green (HLA-A2).

Similar to LLL-specific TCRs, TCRs specific for the HLA-A*02:01-restricted spike epitope YLQ lack structural diversity^{15,16,41}. The YLQ mutation with the highest frequency (0.56%) is P272L at TCR-contacting position P4^{40,54}. This variant was not recognized by >175 individual YLQ-specific TCRs isolated from COVID-19 CPs and vaccinees⁴¹, suggesting that P272L evades T cell responses. Moreover, the P272L mutation arose in >100 different SARS-CoV-2 lineages, including VOCs, indicating transmission⁴¹. Interestingly, the majority (~85%) of HLA-A*02:01-restricted TCRs specific for the YLQ spike epitope, which is unrelated in sequence to the nucleocapsid LLL epitope, also use the TRAV12-1 or TRAV12-2 gene segments^{15,41}. However, the α chains of LLL6E and LLL8 are displaced by ~4.5 Å towards the N-terminus of the LLL peptide compared to their position in all four TCR-YLQ-HLA-A2 complex structures³⁸⁻⁴¹. This displacement, which is likely imposed by the different peptides these TCRs recognize and/or by the different β chains they utilize, results in a different set of contacts between the CDR1 α and CDR2 α loops and HLA-A2. As noted previously for the LLL8 complex structure²⁸, several TCR-pMHC complex structures containing TRAV12-2 TCRs engaging HLA-A2 with different peptides exhibit a shared α chain recognition mode, including the CDR1 α Gln residue hydrogen bonding with the peptide backbone (as seen with LLL6 and LLL8). This highlights a favorable interaction motif likely to be observed in many, but not all, other TCR-pMHC interactions containing TRAV12-2 and HLA-A2.

Modeling of these complexes with deep learning methods TCRmodel2 (based on AlphaFold2) and AlphaFold3 extends our recent findings on modeling accuracy for unseen TCR-pMHC complexes⁵³. While high accuracy models were generated for two complexes by one or both AlphaFold methods, the other complex from this study was not modeled accurately by either method based on top-ranked model, and even considering the full sets of 1000 models, no high accuracy

models were generated for that complex. Additionally, the fact that relatively high confidence scores were output for inaccurate or less accurate models indicates that improved scoring is needed, as also noted recently for antibody-antigen complex modeling using AlphaFold in the Critical Assessment of Predicted Interactions (CAPRI) experiment^{76,77}. Even with extra sampling (1000 seeds per complex, corresponding to 5000 models per complex), the AlphaFold3 antibody-antigen success for high accuracy complexes (DockQ > 0.8) was reported to be less than 30% by the DeepMind team⁶². While AlphaFold represents a major improvement over previous methods for structure prediction of TCR and antibody complexes, additional approaches, such as optimizations of AlphaFold3 (or AlphaFold2) and the development of other novel methods, will be needed to consistently model immune recognition with high accuracy.

The high AlphaFold modeling accuracy observed for the Q04-SPR-HLA-B7 complex in this study, while anecdotal, provides an example of a predictive success for AlphaFold2 and AlphaFold3, but raises the question of whether the PDB training sets of those deep learning models contained any highly related TCR-pMHC complex structures. Searches of the Q04 V α and V β sequences against TCR V domain structures from the PDB in TCR3d⁴⁶ did not identify any high identity hits to both TCR chains, while hits with up to 59% identity and 94% identity were observed for the individual V α and V β sequences, respectively. Additionally, no TCR-pMHC complexes containing SPR-HLA-B7 were available in the PDB prior to the September 2021 training date cutoff, or prior to the structures reported in this study. Therefore, no PDB structures would have enabled AlphaFold to have clear training examples of interactions containing the epitope interacting with a closely related TCR.

By adding to the small (eight structures) but growing dataset of TCR-pMHC complex structures with SARS-CoV-2 epitope recognition,

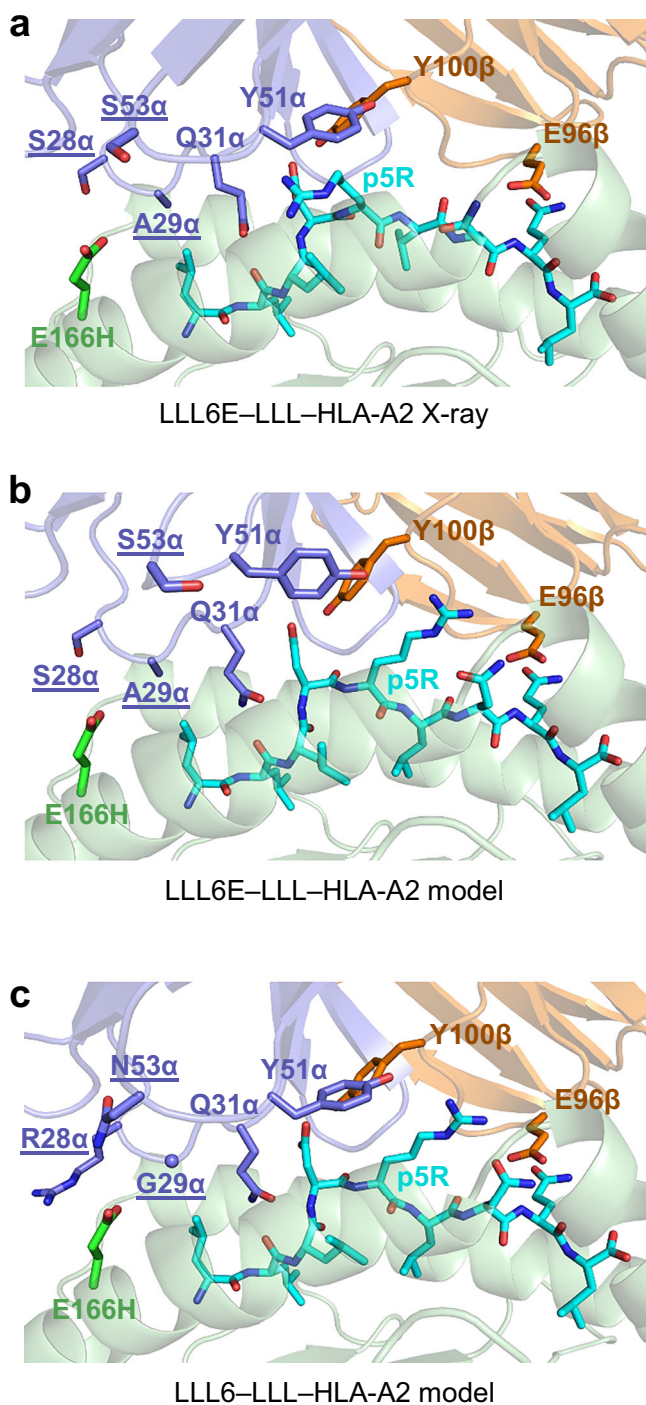


Fig. 8 | Interface of modeled of LLL6 TCR complex and comparison between germline genes. The LLL6E-LLL-HLA-A2 X-ray structure (a) is shown in comparison with corresponding modeled complex (b), as well as modeled complex with native LLL6, containing TRAV12-2 instead of TRAV12-1 α chain residues (c). Peptides are shown as cyan sticks, TCR chains are shown as blue (α chain) and orange (β chain) cartoon, and MHC is shown as green cartoon, with selected TCR and MHC interface residues shown as sticks and labeled. Labels for TCR α chain residues that differ between TRAV12-1 and TRAV12-2 are underlined. Both models were generated by AlphaFold3.

these new structures enable a better understanding of immune diversity and viral escape for SARS-CoV-2 and other variable viruses. These can potentially inform next-generation coronavirus vaccine design, as well as improved structural modeling strategies that can

complement current experimental structural characterization approaches.

Methods

Protein preparation

The isolation and characterization of SPR- and LLL-specific TCRs from COVID-19 CPs was described previously^{28–30}. Soluble TCRs Q04, CLB, LLL6, and LLL6E for affinity measurements and structure determinations were produced by in vitro folding from inclusion bodies expressed in *Escherichia coli*. Codon-optimized genes encoding the α and β chains of these TCRs (TCR Q04 residues 1–206 and 1–243; TCR CLB residues 1–204 and 1–243; LLL6 residues 1–204 and 1–243; TCR LLL6E residues 1–203 and 1–243, respectively) were synthesized (Supplementary Table 13) and cloned into the expression vector pET22b (GenScript). An interchain disulfide (C α Cys160–C β Cys170 in Q04; C α Cys158–C β Cys170 in CLB; C α Cys158–C β Cys170 in LLL6; C α Cys157–C β Cys170 in LLL6E) was engineered to increase the folding yield of TCR $\alpha\beta$ heterodimers. The mutated α and β chains were expressed separately as inclusion bodies in BL21(DE3) *E. coli* cells (Agilent Technologies). Bacteria were grown at 37 °C in LB medium to OD₆₀₀ = 0.6–0.8 and induced with 1 mM isopropyl- β -D-thiogalactoside. After incubation for 3 h, the bacteria were harvested by centrifugation and resuspended in 50 mM Tris-HCl (pH 8.0) containing 0.1 M NaCl and 2 mM EDTA. Cells were disrupted by sonication. Inclusion bodies were washed with 50 mM Tris-HCl (pH 8.0) and 5% (v/v) Triton X-100, then dissolved in 8 M urea, 50 mM Tris-HCl (pH 8.0), 10 mM EDTA, and 10 mM DTT. For in vitro folding, the TCR α (45 mg) and β (35 mg) chains were mixed and diluted into 1 liter folding buffer containing 5 M urea, 0.4 M L-arginine-HCl, 100 mM Tris-HCl (pH 8.0), 3.7 mM cystamine, and 6.6 mM cysteamine. After dialysis against 10 mM Tris-HCl (pH 8.0) for 72 h at 4 °C, the folding mixture was concentrated 20-fold and dialyzed against 50 mM MES buffer (pH 6.0). After removal of the precipitate formed at pH 6.0 by centrifugation, the supernatant was dialyzed overnight at 4 °C against 20 mM Tris-HCl (pH 8.0), 20 mM NaCl. Disulfide-linked Q04, CLB, and LLL6 TCR heterodimers were purified using consecutive Superdex 200 (20 mM Tris-HCl (pH 8.0), 20 mM NaCl) and Mono Q (20 mM Tris-HCl (pH 8.0), 0–1.0 M NaCl gradient) FPLC columns (GE Healthcare).

Soluble HLA-B7 loaded with SPR peptide (SPRWYFYLL) and HLA-A2 loaded with LLL peptide (LLLDRLNQL) peptide were prepared by in vitro folding of *E. coli* inclusion bodies as described⁵². Correctly folded SPR-HLA-B7, LLL-HLA-A2 complexes were purified using sequential Superdex 200 (20 mM Tris-HCl (pH 8.0), 20 mM NaCl) and Mono Q columns (20 mM Tris-HCl (pH 8.0), 0–1.0 M NaCl gradient). To produce biotinylated HLA-B7 and HLA-A2, a C-terminal tag (GGGLNDIFEAQKIEWHE) was attached to the HLA-B*07:02 and HLA-A*02:01 heavy chain, respectively. Biotinylation was carried out with BirA biotin ligase (Avidity).

Crystallization and data collection

For crystallization of TCR-pMHC complexes, TCRs Q04 and CLB were mixed with SPR-HLA-B7 and TCR LLL6E was mixed with LLL-HLA-A2 in a ratio of 1:1 and concentrated to 10 mg/ml. Crystals were obtained at room temperature by vapor diffusion in hanging drops. The Q04-SPR-HLA-B7 complex crystallized in 0.1 M Tris-HCl (pH 8.5), 0.2 M calcium chloride, and 13% (w/v) PEG 3350. Crystals of the CLB-SPR-HLA-B7 complex grew in 0.1 M ammonium citrate dibasic and 20% (w/v) PEG 3350. Crystals of LLL6E-LLL-HLA-A2 were obtained in 0.2 M potassium sodium tartrate tetrahydrate and 20% (w/v) PEG 3350. Crystals of SPR-HLA-B7 grew in 0.1 M HEPES (pH 7.5), 0.2 M ammonium acetate, and 24% (w/v) PEG 3350. All crystals were cryoprotected with 20% (w/v) glycerol and flash-cooled. X-ray diffraction data were collected at beamline BL19U1 and BL02U1 of the National Facility for Protein Science in Shanghai (NFPS), Shanghai Synchrotron Radiation Facility. Diffraction data were indexed, integrated, and

scaled using the program XDS⁷⁸. Data collection statistics are shown in Supplementary Table 2.

Structure determination and refinement

Before structure determination and refinement, all data reductions were performed using the CCP4 software suite⁷⁹. Structures were determined by molecular replacement with the program Phaser⁸⁰ and refined with Phenix⁸¹. The models were further refined by manual model building with Coot⁸² based on $2F_o - F_c$ and $F_o - F_c$ maps. SPR-HLA-B7 (7LGD)²⁹ and AlphaFold-modeled TCR Q04 were used as search models to determine the orientation and position of the Q04-SPR-HLA-B7 complex. The orientation and position parameters of unbound SPR-HLA-B7 were obtained using the corresponding component of the CLB-SPR-HLA-B7 complex as a search model. The α chain of TCR TID3 (6DFX)⁸³, the β chain of TCR DN6 (4ONH)⁸⁴, and SPR-HLA-B7 (7LGD)²⁹ with the CDRs removed were used as search models for molecular replacement to determine the structure of the CLB-SPR-HLA-B7 complex. The α chain of TCR 12-6 (6VRM)⁵², the β chain of TCR AS8.4 (8CX4)⁸⁵, and LLL-HLA-A2 (7KGQ)⁵⁶ with the CDRs removed, were used as search models to determine the orientation and position of the LLL6E-LLL-HLA-A2 complex. Refinement statistics are summarized in Supplementary Table 2. Contact residues were identified with the CONTACT program⁷⁹ and were defined as residues containing an atom 4.0 Å or less from a residue of the binding partner. The PyMOL program (<https://pymol.org/>) was used to prepare figures.

Surface plasmon resonance analysis

The interaction of TCRs Q04, CLB, LLL6, and LLL6E with pMHC was assessed by surface plasmon resonance using a Biacore 8 K biosensor at 25 °C. Biotinylated SPR-HLA-B7 or LLL-HLA-A2 ligand was immobilized on a streptavidin-coated Biacore SA chip (GE Healthcare) at around 200 and 1000 resonance units (RU), respectively. The remaining streptavidin sites were blocked with 20 μ M biotin solution. An additional flow cell was injected with free biotin alone to serve as a blank control. For analysis of TCR binding, solutions containing different concentrations of Q04, CLB, LLL6, or LLL6E were flowed sequentially over chips immobilized with SPR-HLA-B7 or LLL-HLA-A2 ligand, or the blank. Dissociation constants (K_D s) were calculated by fitting equilibrium and kinetic data to a 1:1 binding model using BIA Evaluation 3.1 and Biacore Insight Evaluation software.

Computational sequence and structural analysis

SPR and LLL epitope variant frequencies were calculated from nucleocapsid protein sequences in the GISAID database (www.gisaid.org)⁵⁴, downloaded in February 2025. Frequencies are from a total of 16,931,861 nucleocapsid protein sequences present in the database. Representative nucleocapsid protein sequences for other coronaviruses were obtained from NCBI and aligned using MAFFT software⁸⁶ to generate a multiple sequence alignment, which was used to obtain sequences corresponding to the SPR and LLL epitope positions in those viruses. Reference PDB structures of Class I TCR-pMHC complexes were obtained from the TCR3d database⁴⁶, removing redundant complexes that contained the same TCR or engineered variants of a TCR, and four complexes with reverse polarity TCR binding were also omitted. This resulted in a total of 130 Class I reference complex structures. TCR-pMHC docking angle calculations were performed on the TCR3d site. Prediction of Q04, CLB, LLL6E, and LLL8 TCR binding effects ($\Delta\Delta G$ s) for epitope variants and orthologs was performed using computational mutagenesis in Rosetta (v.2.3)⁵¹, which we previously used to predict TCR-pMHC affinity changes⁸⁷. As in the previous study, off-rotamer side chain minimization was enabled before and after substitution (command line flags: “-min_interface -min_chi”). For the LLL8 TCR-pMHC complex, due to its moderately lower resolution (3.18 Å), we pre-processed the structure using the

Rosetta FastRelax protocol⁸⁸ (Rosetta v3.5) to perform constrained local backbone and side chain minimization prior to computational mutagenesis.

Structural modeling and model assessment

TCRmodel2⁵⁹ and AlphaFold3⁶² were used to predict the structures of the TCR-pMHC complexes from sequence. The modeling included the variable domains of the TCR α and β chains, and the $\alpha 1$ and $\alpha 2$ domains of the MHC. Both methods were run on a local computer cluster using 200 seeds to generate a total of 1000 models for each complex. AlphaFold3 was downloaded from its official GitHub repository in December 2024. A maximum template date cutoff of September 30, 2021 was set for both TCRmodel2 and AlphaFold3, such that only PDB structures released on or before that date were permitted as templates, and the AlphaFold2.3 (used by TCRmodel2) and AlphaFold3 models were both trained using PDB structures with a release date cutoff of September 30, 2021^{61,62}. The models for each complex were ranked using AlphaFold2's multimer model confidence score ($0.8 \times \text{ipTM} + 0.2 \times \text{pTM}$)⁶⁰, the default ranking metric in TCRmodel2, and the top-ranked model for each complex was selected for accuracy assessments. This confidence score differs slightly from the AlphaFold3 ranking score⁵⁹, which additionally includes terms for disorder and clashes, but the AlphaFold2 model confidence score formulation was used for consistency of model ranking in this study. Interface pLDDT (I-pLDDT) values were calculated by averaging pLDDT values for all residues within 4 Å of the binding interface in the modeled complex. For AlphaFold3 models, which have separate pLDDT values for each atom, residue pLDDT values were obtained by averaging atom pLDDT values for each residue. Model accuracy values, including interface RMSD, ligand RMSD, DockQ score, and CAPRI accuracy level were calculated using the DockQ program⁸⁹ by comparing the modeled complex with the corresponding X-ray structure.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Atomic coordinates and structure factors have been deposited in the Protein Data Bank under accession codes 9WBD (Q04-SPR-HLA-B7), 9J4T (CLB-SPR-HLA-B7), 9J4U (LLL6E-LLL-HLA-A2), and 9J4V (SPR-HLA-B7).

References

1. Phelan, A. L., Katz, R. & Gostin, L. O. The novel coronavirus originating in Wuhan, China: challenges for global health governance. *JAMA* **323**, 709–710 (2019).
2. Wu, F. et al. A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265–269 (2020).
3. Zhu, N. et al. A novel coronavirus from patients with pneumonia in China, 2019. *N. Engl. J. Med.* **382**, 727–733 (2020).
4. Jeyanathan, M. et al. Immunological considerations for COVID-19 vaccine strategies. *Nat. Rev. Immunol.* **20**, 615–632 (2020).
5. Rydzynski, M., Moderbacher, C. et al. Antigen-specific adaptive immunity to SARS-CoV-2 in acute COVID-19 and associations with age and disease severity. *Cell* **183**, 996–1012 (2020).
6. Sette, A. & Crotty, S. Adaptive immunity to SARS-CoV-2 and COVID-19. *Cell* **184**, 861–880 (2021).
7. Stadler, E. et al. Determinants of passive antibody efficacy in SARS-CoV-2 infection: a systematic review and meta-analysis. *Lancet Microbe* **4**, e883–e892 (2023).
8. Robbiani, D. F. et al. Convergent antibody responses to SARS-CoV-2 in convalescent individuals. *Nature* **584**, 437–442 (2020).
9. Moss, P. The T cell immune response against SARS-CoV-2. *Nat. Immunol.* **23**, 186–193 (2022).

10. Vardhana, S., Baldo, L., Morice, W. G. 2nd & Wherry, E. J. Understanding T cell responses to COVID-19 is essential for informing public health strategies. *Sci. Immunol.* **7**, eabo1303 (2022).
11. Kent, S. J. et al. Disentangling the relative importance of T cell responses in COVID-19: leading actors or supporting cast?. *Nat. Rev. Immunol.* **22**, 387–397 (2022).
12. Ramirez, S. I. et al. Early antiviral CD4+ and CD8+ T cells are associated with upper airway clearance of SARS-CoV-2. *JCI Insight* **9**, e186078 (2024).
13. Nelde, A. et al. SARS-CoV-2-derived peptides define heterologous and COVID-19-induced T cell recognition. *Nat. Immunol.* **22**, 74–85 (2020).
14. Sekine, T. et al. Robust T cell immunity in convalescent individuals with asymptomatic or mild COVID-19. *Cell* **183**, 158–168 (2020).
15. Shomuradova, A. S. et al. SARS-CoV-2 epitopes are recognized by a public and diverse repertoire of human T cell receptors. *Immunity* **53**, 1245–1257 (2020).
16. Schulien, I. et al. Characterization of pre-existing and induced SARS-CoV-2-specific CD8+ T cells. *Nat. Med.* **27**, 78–85 (2021).
17. Molodtsov, I. A. et al. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)-specific T cells and antibodies in coronavirus disease 2019 (COVID-19) protection: a prospective study. *Clin. Infect. Dis.* **75**, e1–e9 (2022).
18. Naranbhai, V. et al. T cell reactivity to the SARS-CoV-2 Omicron variant is preserved in most but not all individuals. *Cell* **185**, 1259 (2022).
19. Ng, O. W. et al. Memory T cell responses targeting the SARS coronavirus persist up to 11 years post-infection. *Vaccine* **34**, 2008–2014 (2016).
20. Grifoni, A. et al. SARS-CoV-2 human T cell epitopes: adaptive immune responses against COVID-19. *Cell Host Microbe* **29**, 1076–1092 (2021).
21. Wu, W., Cheng, Y., Zhou, H., Sun, C. & Zhang, S. The SARS-CoV-2 nucleocapsid protein: its role in the viral life cycle, structure and functions, and use as a potential target in the development of vaccines and diagnostics. *Viol. J.* **20**, 6 (2023).
22. Maghsood, F. et al. SARS-CoV-2 nucleocapsid: biological functions and implication for disease diagnosis and vaccine design. *Rev. Med. Virol.* **33**, e2431 (2023).
23. Dangi, T., Class, J., Palacio, N., Richner, J. M. & Penaloza MacMaster, P. Combining spike- and nucleocapsid-based vaccines improves distal control of SARS-CoV-2. *Cell Rep.* **36**, 109664 (2021).
24. López-Muñoz, A. D., Kosik, I., Holly, J. & Yewdell, J. W. Cell surface SARS-CoV-2 nucleocapsid protein modulates innate and adaptive immunity. *Sci. Adv.* **8**, eabp9770 (2022).
25. Zhang, B. et al. Comparing the nucleocapsid proteins of human coronaviruses: structure, immunoregulation, vaccine, and targeted drug. *Front. Mol. Biosci.* **9**, 761173 (2022).
26. Yu, H., Guan, F., Miller, H., Lei, J. & Liu, C. The role of SARS-CoV-2 nucleocapsid protein in antiviral immunity and vaccine development. *Emerg. Microbes Infect.* **12**, e2164219 (2023).
27. Le Bert, N. et al. SARS-CoV-2-specific T cell immunity in cases of COVID-19 and SARS, and uninfected controls. *Nature* **584**, 457–462 (2020).
28. Choy, C. et al. SARS-CoV-2 infection establishes a stable and age-independent CD8+ T cell response against a dominant nucleocapsid epitope using restricted T cell receptors. *Nat. Commun.* **14**, 6725 (2023).
29. Lineburg, K. E. et al. CD8+ T cells specific for an immunodominant SARS-CoV-2 nucleocapsid epitope cross-react with selective seasonal coronaviruses. *Immunity* **54**, 1055–1065 (2021).
30. Nguyen, T. H. O. et al. CD8+ T cells specific for an immunodominant SARS-CoV-2 nucleocapsid epitope display high naive precursor frequency and TCR promiscuity. *Immunity* **54**, 1066–1082 (2021).
31. Peng, Y. et al. An immunodominant NP₁₀₅₋₁₁₃-B*07:02 cytotoxic T cell response controls viral replication and is associated with less severe COVID-19 disease. *Nat. Immunol.* **23**, 50–61 (2022).
32. Heitmann, J. S. et al. A COVID-19 peptide vaccine for the induction of SARS-CoV-2 T cell immunity. *Nature* **601**, 617–622 (2022).
33. Heitmann, J. S. et al. Phase I/II trial of a peptide-based COVID-19 T-cell activator in patients with B-cell deficiency. *Nat. Commun.* **14**, 5032 (2023).
34. Arieta, C. M. et al. The T-cell-directed vaccine BNT162b4 encoding conserved non-spike antigens protects animals from severe SARS-CoV-2 infection. *Cell* **186**, 2392–2409 (2023).
35. Peng, Y. et al. Broad and strong memory CD4+ and CD8+ T cells induced by SARS-CoV-2 in UK convalescent individuals following COVID-19. *Nat. Immunol.* **21**, 1336–1345 (2020).
36. Ferretti, A. P. et al. Unbiased screens show CD8+ T cells of COVID-19 patients recognize shared epitopes in SARS-CoV-2 that largely reside outside the spike protein. *Immunity* **53**, 1095–1107 (2020).
37. Meyer-Olson, D. et al. Limited T cell receptor diversity of HCV-specific T cell responses is associated with CTL escape. *J. Exp. Med.* **200**, 307–319 (2004).
38. Szeto, C. et al. Molecular basis of a dominant SARS-CoV-2 spike-derived epitope presented by HLA-A*02:01 recognised by a public TCR. *Cells* **10**, 2646 (2021).
39. Chaurasia, P. et al. Structural basis of biased T cell receptor recognition of an immunodominant HLA-A2 epitope of the SARS-CoV-2 spike protein. *J. Biol. Chem.* **297**, 101065 (2021).
40. Wu, D. et al. Structural assessment of HLA-A2-restricted SARS-CoV-2 spike epitopes recognized by public and private T-cell receptors. *Nat. Commun.* **13**, 19 (2022).
41. Dolton, G. et al. Emergence of immune escape at dominant SARS-CoV-2 killer T cell epitope. *Cell* **185**, 2936–2951 (2022).
42. Wu, D., Efimov, G. A., Bogolyubova, A. V., Pierce, B. G. & Mariuzza, R. A. Structural insights into protection against a SARS-CoV-2 spike variant by T cell receptor diversity. *J. Biol. Chem.* **299**, 103035 (2023).
43. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
44. Rudolph, M. G., Stanfield, R. L. & Wilson, I. A. How TCRs bind MHCs, peptides, and coreceptors. *Annu. Rev. Immunol.* **24**, 419–466 (2006).
45. Pierce, B. G. & Weng, Z. A flexible docking approach for prediction of T cell receptor-peptide-MHC complexes. *Protein Sci.* **22**, 35–46 (2013).
46. Lin, V. et al. TCR3d 2.0: expanding the T cell receptor structure database with new structures, tools and interactions. *Nucleic Acids Res.* **53**, D604–D608 (2025).
47. Lawrence, M. C. & Colman, P. M. Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* **234**, 946–950 (1993).
48. Wodak, S. J. & Janin, J. Structural basis of macromolecular recognition. *Adv. Protein Chem.* **61**, 9–73 (2002).
49. Feng, D., Bond, C. J., Ely, L. K., Maynard, J. & Garcia, K. C. Structural evidence for a germline-encoded T cell receptor-major histocompatibility complex interaction ‘codon’. *Nat. Immunol.* **8**, 975–983 (2007).
50. Marrack, P., Scott-Browne, J. P., Dai, S., Gapin, L. & Kappler, J. W. Evolutionarily conserved amino acids that control TCR-MHC interaction. *Annu. Rev. Immunol.* **26**, 171–203 (2008).
51. Kortemme, T., Kim, D. E. & Baker, D. Computational alanine scanning of protein-protein interfaces. *Sci. STKE* **2004**, pl2 (2004).
52. Wu, D., Gallagher, D. T., Gowthaman, R., Pierce, B. G. & Mariuzza, R. A. Structural basis for oligoclonal T cell recognition of a shared p53 cancer neoantigen. *Nat. Commun.* **11**, 2908 (2020).

53. Wu, D. et al. Structural characterization and AlphaFold modeling of human T cell receptor recognition of NRAS cancer neoantigens. *Sci. Adv.* **10**, eadq6150 (2024).
54. Elbe, S. & Buckland-Merrett, G. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob. Chall.* **1**, 33–46 (2017).
55. Tian, J. et al. T cell immune evasion by SARS-CoV-2 JN.1 escapees targeting two cytotoxic T cell epitope hotspots. *Nat. Immunol.* **26**, 265–278 (2025).
56. Szeto, C. et al. The presentation of SARS-CoV-2 peptides by the common HLA-A*02:01 molecule. *iScience* **24**, 102096 (2021).
57. Yin, R., Feng, B. Y., Varshney, A. & Pierce, B. G. Benchmarking AlphaFold for protein complex modeling reveals accuracy determinants. *Protein Sci.* **31**, e4379 (2022).
58. Yin, R. & Pierce, B. G. Evaluation of AlphaFold antibody-antigen modeling with implications for improving predictive accuracy. *Protein Sci.* **33**, e4865 (2024).
59. Yin, R. et al. TCRmodel2: high-resolution modeling of T cell receptor recognition using deep learning. *Nucleic Acids Res.* **51**, W569–W576 (2023).
60. Evans, R. et al. Protein complex prediction with AlphaFold-Multimer. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.10.04.463034> (2021).
61. DeepMind. *AlphaFold v2.3.0 Technical Note* https://github.com/deepmind/alphafold/blob/main/docs/technical_note_v2.3.0.md (2022).
62. Abramson, J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).
63. Lee, E. et al. Identification of SARS-CoV-2 nucleocapsid and spike T-cell epitopes for assessing T-cell immunity. *J. Virol.* **95**, e02002–e02020 (2021).
64. Chiuppesi, F. et al. Synthetic multiantigen MVA vaccine COH04S1 protects against SARS-CoV-2 in Syrian hamsters and non-human primates. *NPJ Vaccines* **7**, 7 (2022).
65. Chiuppesi, F. et al. Safety and immunogenicity of a synthetic multiantigen modified vaccinia virus Ankara-based COVID-19 vaccine (COH04S1): an open-label and randomised, phase 1 trial. *Lancet Microbe* **3**, e252–e264 (2022).
66. Jia, Q. et al. Replicating bacterium-vectored vaccine expressing SARS-CoV-2 membrane and nucleocapsid proteins protects against severe COVID-19-like disease in hamsters. *NPJ Vaccines* **6**, 47 (2021).
67. Matchett, W. E. et al. Cutting edge: nucleocapsid vaccine elicits spike-independent SARS-CoV-2 protective immunity. *J. Immunol.* **207**, 376–379 (2021).
68. Ahn, J. Y. et al. Safety and immunogenicity of two recombinant DNA COVID-19 vaccines containing the coding regions of the spike or spike and nucleocapsid proteins: an interim analysis of two open-label, non-randomised, phase 1 trials in healthy adults. *Lancet Microbe* **3**, e173–e183 (2022).
69. Castro, J. T. et al. Promotion of neutralizing antibody-independent immunity to wild-type and SARS-CoV-2 variants of concern using an RBD-nucleocapsid fusion protein. *Nat. Commun.* **13**, 4831 (2022).
70. Afkhami, S. et al. Respiratory mucosal delivery of next-generation COVID-19 vaccine provides robust protection against both ancestral and variant strains of SARS-CoV-2. *Cell* **185**, 896–915 (2022).
71. LeMaout, J. et al. Age-related dysregulation in CD8 T cell homeostasis: kinetics of a diversity loss. *J. Immunol.* **165**, 2367–2373 (2000).
72. Messaoudi, I. et al. Direct link between MHC polymorphism, T cell avidity, and diversity in immune defense. *Science* **298**, 1797–1800 (2002).
73. Chen, H. et al. TCR clonotypes modulate the protective effect of HLA class I molecules in HIV-1 infection. *Nat. Immunol.* **13**, 691–700 (2012).
74. Price, D. A. et al. Public clonotype usage identifies protective Gag-specific CD8+ T cell responses in SIV infection. *J. Exp. Med.* **206**, 923–936 (2009).
75. Gaiha, G. D. et al. Structural topology defines protective CD8+ T cell epitopes in the HIV proteome. *Science* **364**, 480–484 (2019).
76. Raouraoua, N., Lensink, M. F. & Brysbaert, G. Massive sampling strategy for antibody-antigen targets in CAPRI round 55 with MassiveFold. *Proteins* (in press).
77. Gowthaman, R., Park, M., Yin, R., Guest, J. D. & Pierce, B. G. AlphaFold and docking approaches for antibody-antigen and other targets: insights from CAPRI rounds 47–55. *Proteins* (in press).
78. Kabsch, W. XDS. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 125–132 (2010).
79. Collaborative Computational Project No. 4 The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D. Biol. Crystallogr.* **50**, 240–255 (1994).
80. Storoni, L. C., McCoy, A. J. & Read, R. J. Likelihood-enhanced fast rotation functions. *Acta Crystallogr. D. Biol. Crystallogr.* **60**, 432–438 (2004).
81. Afonine, P. V. et al. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D. Biol. Crystallogr.* **68**, 352–367 (2012).
82. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 486–501 (2010).
83. Wang, Y. et al. How C-terminal additions to insulin B-chain fragments create superagonists for T cells in mouse and human type 1 diabetes. *Sci. Immunol.* **4**, eaav7517 (2019).
84. Roy, S. et al. Molecular basis of mycobacterial lipid antigen presentation by CD1c and its recognition by $\alpha\beta$ T cells. *Proc. Natl. Acad. Sci. USA* **111**, E4648–E4657 (2014).
85. Yang, X. et al. Autoimmunity-associated T cell receptors recognize HLA-B*27-bound peptides. *Nature* **612**, 771–777 (2022).
86. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
87. Wu, D., Gowthaman, R., Pierce, B. G. & Mariuzza, R. A. T cell receptors employ diverse strategies to target a p53 cancer neoantigen. *J. Biol. Chem.* **298**, 101684 (2022).
88. Khatib, F. et al. Algorithm discovery by protein folding game players. *Proc. Natl. Acad. Sci. USA* **108**, 18949–18953 (2011).
89. Basu, S. & Wallner, B. DockQ: a quality measure for protein-protein docking models. *PLoS ONE* **11**, e0161879 (2016).

Acknowledgements

This work was supported by National Natural Science Foundation of China Grants 32100985 and 32270995 (to D.W.), by Outstanding Youth Fund of Hunan Provincial Natural Science Foundation Grant 2023JJ10034 (to D.W.), by Science and Technology Innovation Program of Hunan Province Grant 2022RC1209 (to D.W.), by Hunan Provincial Health Commission High-Level Talent Project (to D.W.), by National Institutes of Health Grants GM144083 (to B.G.P.) and AI169181 (to R.A.M.) and by the Intramural Research Program of the National Institutes of Health, National Institute on Aging (to N.-P. Weng). Results in this report are based on work performed at beamlines BL19U1 and BLO2U1 of the National Facility for Protein Science in Shanghai (NFPS), Shanghai Synchrotron Radiation Facility Structural Biology Center. We thank Shuai Zhu, Shuailong Huang and Prof. Xudong Kong (Shanghai Jiao Tong University) for assistance in affinity measurements.

Author contributions

P.Y., G.C., Y.L., X.L., S.S., J.Z., Q.J., H.W., Y.L. and M.E. performed the experiments and data analyses. N.P.W., B.G.P., R.A.M. and D.W. conceived and supervised the project. All authors prepared the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-66322-6>.

Correspondence and requests for materials should be addressed to Brian G. Pierce, Roy A. Mariuzza or Daichao Wu.

Peer review information *Nature Communications* thanks Tao Dong, who co-reviewed with Elie AntounMichael Birnbaum and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025

¹Department of Hepatopancreatobiliary Surgery, The First Affiliated Hospital, Hengyang Medical School, University of South China, Hengyang, Hunan, China. ²Tumor ImmunoMetabolism Institute, Zhuzhou Hospital Affiliated to Xiangya School of Medicine, Central South University, Zhuzhou, Hunan, China. ³School of Chemistry and Chemical Engineering, University of South China, Hengyang, Hunan, China. ⁴W.M. Keck Laboratory for Structural Biology, University of Maryland Institute for Bioscience and Biotechnology Research, Rockville, MD, USA. ⁵Department of Cell Biology and Molecular Genetics, University of Maryland, College Park, MD, USA. ⁶State Key Laboratory of Microbial Metabolism and School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China. ⁷Department of Scientific Research, The First Affiliated Hospital of Jinzhou Medical University, Jinzhou, Liaoning, China. ⁸Laboratory of Molecular Biology and Immunology, National Institute on Aging, National Institutes of Health, Baltimore, MD, USA. ⁹These authors contributed equally: Ping Yuan, Guodong Chen, Yukun Li. ✉e-mail: pierce@umd.edu; rmariuzz@umd.edu; wudaichao@usc.edu.cn