HHS Public Access

Author manuscript

Nature. Author manuscript; available in PMC 2023 September 20.

Published in final edited form as:

Nature. 2022 December ; 612(7941): 771–777. doi:10.1038/s41586-022-05501-7.

Autoimmunity-associated T cell receptors recognize HLA-B*27-bound peptides

Xinbo Yang[1,2,12], **Lee I. Garner**[3,4,12], **Ivan V. Zvyagin**[5,6,12], **Michael A. Paley**[7,12], **Ekaterina A. Komech**[5,6], **Kevin M. Jude**[1,2], **Xiang Zhao**[1,2], **Ricardo A. Fernandes**[1,2], **Lynn M. Hassman**[8], **Grace L. Paley**[8], **Christina S. Savvides**[1,2], **Simon Brackenridge**[3,4], **Max N. Quastel**[3,4], **Dmitriy M. Chudakov**[5,6], **Paul Bowness**[9], **Wayne M. Yokoyama**[7,10,✉], **Andrew J. McMichael**[3,4,✉], **Geraldine M. Gillespie**[3,4,✉], **K. Christopher Garcia**[1,2,11,✉]

[1]Department of Molecular and Cellular Physiology, Stanford University School of Medicine, Stanford, CA, USA.

[2]Department of Structural Biology, Stanford University School of Medicine, Stanford, CA, USA.

[3]NDM Research Building, Nuffield Department of Medicine, University of Oxford, Oxford, UK.

[4]Centre for Immuno-oncology, Nuffield Department of Medicine, University of Oxford, Oxford, UK.

[5]Center for Precision Genome Editing and Genetic Technologies for Biomedicine, Institute of Translational Medicine, Pirogov Russian National Research Medical University, Moscow, Russian Federation.

[6]Genomics of Adaptive Immunity Department, Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Moscow, Russian Federation.

[✉]**Correspondence and requests for materials** should be addressed to Wayne M. Yokoyama, Andrew J. McMichael, Geraldine M. Gillespie or K. Christopher Garcia. yokoyama@dom.wustl.edu; andrew.mcmichael@ndm.ox.ac.uk; geraldine.gillespie@ndm.ox.ac.uk; kcgarcia@stanford.edu.

Author contributions X.Y., L.I.G., G.M.G., A.J.M. and K.C.G. conceived the project and wrote the manuscript. X.Y. conducted yeast display experiments, deep-sequencing data analysis, target prediction, target validation, TCR–pMHC complex purification, crystallization and structural studies. L.I.G. conducted processing of samples from patients with AS, single-cell T cell sequencing analysis and TCR tetramer staining of SCTs. M.A.P. and W.M.Y. conceived and analysed single-cell TCR sequencing from HLA-B*27^{+} AAU and contributed clinical perspectives and manuscript revisions. M.A.P., G.L.P. and L.M.H. identified HLA-B*27^{+} participants with AAU and collected aqueous and blood samples. S.B. designed the SCT staining constructs. M.N.Q. carried out thermal melt analysis. X.Z. carried out independent target validation. R.A.F., X.Y. and C.S.S. developed the algorithm for peptide prediction. K.M.J. conducted structural studies. X.Y. and L.I.G. conducted SPR experiments. E.A.K., I.V.Z. and D.M.C. identified the AS3.1 TCR αβ pair, conducted independent target validation for AS3.1 TCR and provided clinical samples. P.B. provided clinical samples and TCR analysis. W.M.Y., A.J.M., G.M.G. and K.C.G. supervised the project.

Competing interests M.A.P. has received research support from Eli Lilly and Company paid to the university and served as a consultant for AbbVie, Priovant Therapeutics and JK MarketResearch. The remaining authors declare no competing interests.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-022-05501-7.

Peer review information *Nature* thanks Robert Colbert and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

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

Reporting summary

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

Code availability

Custom Perl scripts for the deep-sequencing data processing are available from https://github.com/jlmendozabio/NGSpeptideprepandpred.

[7]Rheumatology Division, Department of Medicine, Washington University School of Medicine, St Louis, MO, USA.

[8]Department of Ophthalmology, Washington University School of Medicine, St Louis, MO, USA.

[9]Nuffield Department of Orthopaedics Rheumatology and Muscuoskeletal Science (NDORMS), Botnar Research Center, University of Oxford, Oxford, UK.

[10]Bursky Center for Human Immunology and Immunotherapy Programs, Washington University School of Medicine, St Louis, MO, USA.

[11]Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, CA, USA.

[12]These authors contributed equally: Xinbo Yang, Lee I. Garner, Ivan V. Zvyagin, Michael A. Paley.

Abstract

Human leucocyte antigen B*27 (HLA-B*27) is strongly associated with inflammatory diseases of the spine and pelvis (for example, ankylosing spondylitis (AS)) and the eye (that is, acute anterior uveitis (AAU))[1]. How HLA-B*27 facilitates disease remains unknown, but one possible mechanism could involve presentation of pathogenic peptides to $CD8^+$ T cells. Here we isolated orphan T cell receptors (TCRs) expressing a disease-associated public β-chain variable region–complementary-determining region 3β (BV9–CDR3β) motif[2–4] from blood and synovial fluid T cells from individuals with AS and from the eye in individuals with AAU. These TCRs showed consistent α-chain variable region (AV21) chain pairing and were clonally expanded in the joint and eye. We used HLA-B*27:05 yeast display peptide libraries to identify shared self-peptides and microbial peptides that activated the AS-and AAU-derived TCRs. Structural analysis revealed that TCR cross-reactivity for peptide–MHC was rooted in a shared binding motif present in both self-antigens and microbial antigens that engages the BV9–CDR3β TCRs. These findings support the hypothesis that microbial antigens and self-antigens could play a pathogenic role in HLA-B*27-associated disease.

HLA-B*27 is strongly associated with AS[5], AAU[6] and reactive arthritis (ReA)[7]. Whereas AS is characterized by chronic inflammation of the spine, sacroiliac and sometimes peripheral joints, AAU involves episodic inflammation of the iris and ciliary body in the eye[8]. Although AS and AAU can occur in isolation, more than half of patients with AS experience at least one episode of AAU[9]. ReA has a similar clinical phenotype to AS but is distinguished by onset following bacterial infection, typically with *Chlamydia*, *Klebsiella*, *Salmonella*, *Shigella* or *Yersinia*[10,11], suggesting a role for these microorganisms in ReA and possibly in AS and AAU.

The pathogenic mechanisms that link HLA-B*27 to AS and AAU are not fully understood. The arthritogenic peptide hypothesis[12] proposes that CD8+ T cells, primed by microbial peptides presented by HLA-B*27, subsequently interact with HLA-B*27-bound self-peptides. However, until now, methods to identify candidate peptides have been inadequate. Animal models of HLA-B*27-associated disease require superphysiological HLA-B*27 and human β_2-microglobulin (β_2m) transgenic insertions[13,14] and do not reproduce AAU[15]. Other disease models describe the propensity of the HLA-B*27 heavy chain to dimerize

or aggregate, triggering the unfolded protein response or innate receptor engagement, and involve type I (tumour necrosis factor) and interleukin (IL)-17–IL-23 inflammatory cytokine axes[16–19].

Studies of independent patient cohorts have reported shared (public) TCRs utilizing Vβ9 (TRBV9)–CDR3–Jβ2.3 (TRBJ2.3) chains in blood $CD8^+$ T cells from HLA-B*27^+ individuals with AS, but not healthy HLA-B*27^+ control participants, with further enrichment in the joint[2–4,20–23]. Genome-wide association studies have highlighted strong genetic links with endoplasmic reticulum aminopeptidase 1 (*ERAP1*) single-nucleotide polymorphisms in HLA-B*27:05^+ patients with AS[24]. ERAP1 participates in amino-terminal trimming of major histocompatibility complex (MHC) class I-restricted peptides[25,26], and *ERAP1* polymorphisms affect peptide repertoire generation[27]. Finally, certain HLA-B*27 subtypes are not associated with AS[28]. These findings necessitate reappraisal of the arthritogenic peptide hypothesis, focusing on disease-associated TCRs.

Here, by using TCR-driven selection of HLA-B*27:05-based yeast display libraries, we identified self-antigens and microbial antigens recognized by the reported public $TRBV9^+$ $CD8^+$ TCRs derived from the synovial fluid and blood of HLA-B*27^+ patients with AS and from the aqueous humour and blood of HLA-B*27^+ patients with AAU.

Single-cell sequencing of AS and AAU T cells

Enrichment of TRBV9–CDR3–TRBJ2.3 sequence motifs in patients with AS compared to HLA-B*27^+ healthy individuals was previously described[2–4], although TCRα partners were not reported. From single-cell RNA sequencing (scRNA-seq) of peripheral blood mononuclear cell (PMBC) samples from HLA-B*27^+ patients with AS (Fig. 1a), we identified $TRBV9^+$ T cells expressing the AS-enriched CDR3 Y/FSTDTQ–TRBJ2.3 motif and their corresponding TCRα chains (Extended Data Fig. 1b). Owing to low reported frequencies (about 1 in 10^5 of αβ T cells[3]; Fig. 1b and Extended Data Fig. 1c), we pre-enriched for $TRBV9^+CD8^+$ T cells (Fig. 1a (group 1) and Extended Data Fig. 1a (group 1.1)), recombining these with various ratios of sample-specific $CD8^+BV9^-$ cells (Extended Data Fig. 1d,e and Supplementary Table 1), a strategy adopted because a non-BV9-linked VGLY motif was previously identified in ReA[22]. This approach retrieved a TRBV9–CDR3 VGLY sequence and its corresponding (AS2 TCR) TRAV21 chain in one patient, AS1541 (Fig. 1b and Supplementary Table 1). Subsequently, we analysed four additional patient samples sorted exclusively for $CD3^+CD8^+TRBV9^+$ (10,000 cells per library; Extended Data Fig. 1a (group 1.2)). Patient AS1455 yielded 11 barcodes (of about 10,000 barcodes total) with BV9–CDR3 clonotypes described previously, including four VGLYSTDTQ (AS4.1 TCR), one VGLFSTDTQ (AS4.2 TCR) and one VATYSTDTQ barcode (AS4.3 TCR[2–4]; Fig. 1b and Extended Data Fig. 1f). Five additional barcodes covering three clonotypes had the canonical CDR3β, albeit with different TRBJ gene segment usage and varying CDR3α (Supplementary Table 1). These TCRs used the TRAV21 chain with no obvious CDR3α patterns. Other $BV9^+$ expansions that have not been reported to be enriched in disease (samples AS1311 (about 15%) and AS1803 (about 70%)) were not selected for follow-up[2,4].

The AS-related Y/FSTDTQ-BJ2.3 motif is enriched in synovial fluid compared to blood[4]. We interrogated four synovial fluid-derived samples (Fig. 1a (group 2)). One TCR αβ pair (AS3.1, patient P (AS)), inferred by bulk TCR sequencing with the VGLYSTDTQ CDR3 motif and the TRAV21 chain found in about 30% of all TCRαβ reads, respectively (Extended Data Fig. 2a), was subsequently confirmed by TCR scRNA-seq (Supplementary Table 2). Y/FSTDTQ CDR3 motifs were also enriched in synovial fluid from other patients with AS (patients K, N and S; Fig. 1b). Bulk TCR sequencing of paired samples confirmed at least 10-fold enrichment of the BV9-Y/FSTDTQ-BJ2.3 motif in synovial fluid over blood (Extended Data Fig. 2b).

An unbiased approach identified expanded clonotypes in HLA-B*27^{+} patients with AAU. Paired eye and blood samples from four patients underwent TCR scRNA-seq (Fig. 1a (group 3)). One patient (UV027) had AS, whereas UV019 and UV180 had no extra-ocular disease. Each ocular sample contained at least one TCR sequence from more than ten barcodes, consistent with clonal expansion (Fig. 1c). Expansion of TCRs was 10- to 100-fold greater in the eye than in the blood, suggesting eye-specific recruitment or expansion (Fig. 1c). Notably, expanded clonotypes from patient UV180 (AU1.1 and AU1.2) matched the AS-derived TRBV9-Y/FSTDTQ-BJ2.3 paired with the TRAV21 chain and distinct CDR3α sequences (Fig. 1d), even though UV180 had no clinical evidence of AS. Three additional unexpanded clonotypes with the BV9-Y/FSTDTQ-BJ2.3 motif paired with TRAV21 were found: one from UV180 (AU1.3) and two from UV027b (AU2.1 and AU2.2; Fig. 1d). Collectively, these data demonstrate uniform pairing of the BV9-Y/FSTDTQ-BJ2.3 motif with the TRAV21 chain in HLA-B*27^{+} patients with AS and/or AAU.

TCR screening and antigen prediction

We used the HLA-B*27:05-restricted influenza NP peptide (SRYWAIRTR)-specific TCR (GRb) to validate a stably expressed HLA-B*27:05-based platform encoding three heavy chain amino acid substitutions, HLA-B*27:05$^{(3mut)}$ (outlined in Methods; Extended Data Fig. 3a–h). Five AS-enriched TCRs were expressed as soluble proteins to screen 9-amino acid (AA) and 10-AA HLA-B*27:05$^{(3mut)}$–peptide libraries. AS3.1 TCRs enriched peptides from the 10-AA library alone, whereas AS4 TCRs recovered peptides from both libraries (Fig. 2a and Extended Data Fig. 4a). AS3.1, AS4.2, AS4.3 and AS4.4 TCR tetramers stained fourth-round libraries, indicating TCR-driven selection (Extended Data Fig. 4b,c). Absence of AS4.1 TCR tetramer staining probably reflected very low-affinity interactions between the TCR and peptide–MHC (pMHC).

Deep sequencing of yeast display selections showed sequence convergence by round four. Unique peptides were counted, and positional hotspots were identified (Fig. 2b). AS3.1 TCRs enriched a group of 9-AA peptides within the 10-AA library with similarity to those from AS4 TCR-enriched 9-AA libraries (Fig. 2c), indicating shared peptide recognition. The peptides SRVMLLAPR and ARVMLVAPR were shared by AS3.1, AS4.2, AS4.3 and AS4.2, and AS4.3 and AS4.4 TCRs, respectively (Supplementary Table 3). P4 Met, P5 Leu, P7 Ala and P8 Pro were highly conserved, suggesting a role for TCR recognition (Fig. 2b,c). Despite apparent commonalities, subtle differences were observed. The AS4.3 TCR returned

more diverse peptides, whereas AS4.4 TCR preferentially selected P6 Arg versus P6 Leu, Ile or Val, perhaps owing to differential TRBJ usage.

The 10-AA libraries revealed less obvious convergence of peptide motifs (Extended Data Fig. 4d,e) but mimicked 9-AA library sequences, possibly to retain TCR contact hotspots. This length ambiguity was probably attributable to the Y84A substitution, necessary to accommodate the peptide linker and to facilitate peptide 'sliding'.

The peptides that were enriched following HLA-B*27:05$^{(3mut)}$ yeast display selections provided a peptide-specific recognition landscape for each TCR. We searched Uniprot proteomes of interest (human, *Chlamydia*, *Salmonella*, *Shigella*, *Klebsiella* and *Yersinia* species)[29–31] for possible peptide targets. More than 100 potential peptide targets were shared by at least two TCRs and these were used for in vitro T cell activation studies (Fig. 2d and Supplementary Tables 3 and 4).

Predicted peptide validation

The top 9-AA and 10-AA peptides from each AS TCR selection activated their corresponding AS TCRs in co-culture assays in vitro (Extended Data Fig. 5a–e and Supplementary Table 3), recapitulating the peptide library screening with native HLA-B*27:05. In accordance with its increased cross-reactivity, 17 of 19 synthetic 9-AA peptides were agonists for the AS4.3 TCR. The top 10-AA synthetic peptides recovered for each AS TCR yeast selection (Extended Data Fig. 5f–j and Supplementary Table 3) that activated AS TCRs shared similar motifs to those detected in 9-AA selections.

Stimulation data were combined with deep sequencing to iteratively refine natural peptide prediction. We identified and tested selected peptides derived from human or microbial proteins (Supplementary Table 4). AS4.3 TCR was stimulated by eight human proteome-derived peptides: PRPF3^{self} (TRLALIAPK), JAK3^{self} (DRQQLPAPK), HELQ^{self} (RRVILRAPY), IPO9^{self} (TQMPLVAPV), GLRB^{self} (VQVMLNNPK), RNASEH2B^{self} (GQVMVVAPR), MPP4^{self} (LRQMLQAPH) and SEC14L2^{self} (GRVGDLSPR; Fig. 3d). PRPF3^{self} and GPER1^{self} (GQMWLLAPR) also activated the synovial fluid-derived AS3.1 TCR (Fig. 3a). Additionally, HELQ^{self}, GPER1^{self} and RNASEH2B^{self} peptides stimulated the AS4.2 TCR (Fig. 3c), whereas GPER1^{self} and SEC14L2^{self} only weakly activated the AS4.1 (Fig. 3b) and AS4.4 TCRs (Fig. 3e), respectively.

Predicted peptides from ReA-triggering microorganisms were tested. One peptide, from the UPF0324 inner-membrane protein YEIH^{bac} (LRVMMLAPF), activated AS3.1, AS4.1, AS4.2 and AS4.3 TCRs (Fig. 3a–e). This peptide is conserved across *Escherichia coli*, *Salmonella*, *Shigella* and *Klebsiella* species, and, in *Yersinia*, it harbours a P1 Ile instead of Leu.

We also tested TRBV9^{+} Y/FSTDTQ clones from the synovial fluid of patients with AS (AS8.2, AS8.3, AS8.4, AS8.5, AS9.1 and AS9.2) and ocular clonotypes from patients with AAU (AU1.2 and AU2.1). GPER1^{self}, PRPF3^{self}, RNASEH2B^{self}, YEIH^{bac} and gspD^{bac} elicited broad levels of activation (Fig. 3f–m). AS synovial fluid-derived AS8.4 TCRs showed notably superior activity towards GPER1^{self} and YEIH^{bac} peptides.

We also confirmed that GPER1 and YEIH peptides can be processed in vitro by monitoring CD69 upregulation on SKW-3 T cells expressing AS8.4 in response to HLA-B*27:05-expressing K562 cells transduced with GPER1 and YEIH mini-gene constructs (Extended Data Fig. 6a,b).

To further validate AS-derived TCR recognition of human and microbial peptides, we used TCR tetramers to stain *B2M*-knockout 293T cells transfected with peptide–β_2m–B*2705 single-chain trimers expressing PRPF3^{self}, GPER1^{self}, RNASEH2B^{self}, GLRB^{self} and YEIH^{bac} peptides. For JAK3^{self}, HELQ^{self}, IPO9^{self} or SEC14L2^{self}, which activate AS4.3 TCRs only at a high peptide concentration, no tetramer staining was detected, consistent with low-affinity interactions (Extended Data Fig. 6c).

Differential activation by B*27 subtypes

Certain HLA-B*27 subtypes appear to be protective for AS, including HLA-B*27:09, which differs from HLA-B*27:05 by a single amino acid (D116H)[28]. We pulsed K562-expressing HLA-B*27:05 or HLA-B*27:09 cells with peptides including GPER1^{self}, RNASEH2B^{self}, PRPF3^{self}, MPP4^{self} and YEIH^{bac} and then exposed them to SKW-3 T cells expressing AS4.3, AS8.2, AS8.4, and AU2.1 TCRs. Among the tested peptides, GPER1^{self} (Extended Data Fig. 7a,b), RNASEH2^{self} (Extended Data Fig. 7d,e) and MPP4^{self} (Extended Data Fig. 7m) yielded robust T cell responses (CD69 upregulation) when presented by HLA-B*27:05-expressing K562 cells but not HLA-B*27:09^{+} antigen-presenting cells (APCs). By contrast, comparable T cell activation was observed for PRPF3^{self} and YEIH^{bac} peptides when presented by HLA-B*27:05 or HLA-B*27:09 target cells (Extended Data Fig. 7j,k). Modelled peptide binding based on existing peptide–HLA-B*27:05 and peptide–HLA-B*27:09 crystal structures[32,33] (Protein Data Bank: 2BST, 1JGE and 1K5N) suggested that the most substantial steric hindrance for HLA-B*27:09 should occur with Arg at peptide P9, and also His P9, with its positively charged Nη close to HLA-B*27:09 His116 Nε (Extended Data Fig. 7c,f,n). Peptide P9 Lys or Phe may be better tolerated, potentially explaining the equipotent HLA-B*27:05- and HLA-B*27:09-expressing APCs (Extended Data Fig. 7i,l).

We assessed the thermal stability of HLA-B*27:05 and the disease-protective HLA-B*27:06 and HLA-B*27:09 subtypes plus HLA-B*27:05 (H114Y) (to recapitulate the 114 mutation in the HLA-B*27:05$^{(3mut)}$ construct) refolded with GPER1^{self}, PRPF3^{self} and YEIH^{bac} peptides. All subtypes showed similar thermal stabilities when refolded with the YEIH^{bac} peptide, albeit weakest for HLA-B*27:06. Shared differences were observed for both GPER1^{self} and PRPF3^{self} peptides, with the highest melting temperature (T_m) values observed for HLA*27:05 and HLA-B*27:05(H114Y), followed by HLA-B*27:06 and HLA-B*27:09 (Extended Data Fig. 7o). Overall, our data suggest that features of peptide binding and presentation exhibited by disease-associated versus non-disease-associated HLA-B*27 subtypes could contribute to the pathogenesis of spondyloarthritis[34].

Biophysical and structural analysis

We carried out surface plasmon resonance analysis on the different AS TCR–HLA-B*27:05–peptide interactions (Extended Data Fig. 8a–p). These affinities were 2–12-fold weaker for human peptides than for the bacterial peptide YEIH and are in accord with median effective concentration (EC_{50}) data generated for T cell activation, although the YEIH peptide-based maximum effect (E_{max}) varied for different TCRs (Supplementary Table 5). The equilibrium dissociation constant (K_D) values for TCR–peptide–HLA-B*27:05 interactions were within ranges reported for agonist pMHC Ia-restricted TCRs[35] (Supplementary Table 5). For AS8.4 TCRs, the K_D values recorded for self-peptides (GPER1, 15 μM) and bacterial peptides (YEIH, 1.2 μM) were low (Extended Data Fig. 8o–p), in line with activation data (Fig. 3h).

To visualize how these TCRs mediate broad cross-reactivity, we carried out X-ray crystallographic analysis. Seven AS TCR–HLA-B*27:05 peptide complexes were evaluated, including three AS blood-derived TCRs that recognized the self-peptides $PRPF3^{self}$ (AS3.1, AS4.2, AS4.3) and $RNASEH2B^{self}$ (AS4.3), plus three that recognized the bacterial $YEIH^{bac}$ epitope (AS4.2, AS4.3 and AS8.4; Fig. 4a–d and Extended Data Fig. 9a–h). All TCR CDR footprints (Extended Data Fig. 10a) were superimposable in relation to the mode of peptide–HLA-B*27:05 engagement (Fig. 4e–h and Extended Data Fig. 10b), indicating utilization of a uniform structural solution to achieve optimal pMHC recognition. Tyr98 or Phe98 in the AS TCR CDR3β regions inserted into a pocket formed between the HLA-B*27 α2 helix and Pro at P8 of the peptides (Extended Data Fig. 10c). Tyr98 or Phe98, Ser99 and Thr100 also formed additional contacts with peptide residues P5, P6 and P7 (Fig. 4i,j and Extended Data Fig. 10e–k). Notably, Tyr98 or Phe98, in contrast to Ser99 and Thr100, is not encoded by the TRBJ2–3 germline sequence but is formed after trimming and/or nucleotide addition during V(D)J recombination. This observation supports yeast selection data in which near-exclusive selection of P8 Pro for TCRs that carried the CDR3β Y/FSTDTQ motif was observed. The TRAV21 chain used CDR1α residues (Ala30, Ile31 and Tyr32) to form van der Waals contacts with P4 of the peptide (Fig. 4i,j and Extended Data Fig. 10e–k), similar to a few published TCR–pMHC complexes[36,37]. The TRAV21 CDR3α was minimally involved in peptide recognition, explaining our sequence-disparate CDR3α scRNA-seq data.

TCR contacts to the HLA-B*27:05 heavy chain were mainly (>85%) mediated by the CDR2α, CDR3α and CDR3β regions. Despite their sequence divergence, the various AS TCR CDR3α motifs mediated focused interactions with HLA-B*27:05 α1 helix Gln65 and Lys68. CDR3β Tyr98 or Phe98, Thr100 and CDR2α Gln52, Ser53, Ser54 formed a dense network of hydrogen bonds and van der Waals contacts with the HLA-B*27:05 α2 helix, spanning residues Lys146, Trp147, Ala150, Arg151, Val152, Gln155, Ala158 and Glu163 (Extended Data Fig. 10d). The remaining CDRs interacted minimally with HLA-B*27:05. The steep TCR–pMHC docking angles (25–29°) allowed positioning of the CDRs to facilitate MHC interactions (Extended Data Fig. 10b). In a previous study of AS and ReA, a TRBV23-Y/FSTDTQ-TRBJ2.3β chain rearrangement was also identified along with TRBV9-Y/FSTDTQ-TRBJ2.3 (ref. [38]). Structurally, we predict that TRBV23 CDR1β and CDR2β should also play minor roles in peptide–HLA (pHLA)-B*27:05 interaction.

This contrasts with critical BV9-mediated germline-encoded interactions with HLA-DQ in coeliac disease, in which the CDR1 Leu37 and CDR2 Tyr57 amino acids helped to stabilize the overall complex[39]. Hence, BV9 selection in these HLA-B*27-linked conditions may be driven by non-structural factors relating to convergent and biased V(D)J recombination events[40,41].

Discussion

The recent identification of expanded TRBV9-CDR3β-TRBJ2.3 T cells in HLA-B*27$^+$ patients with AS[2–4] allows reappraisal of the arthritogenic peptide hypothesis. Here we identified the AV21 chain pairings of AS-associated TRBV9-CDR3β-TRBJ2.3 CD8$^+$ T cells from patient PBMCs and synovial fluid samples. Remarkably, TCRs sharing the TRBV9-CDR3β motif with TRAV21 chain pairing were also found in the eye in HLA-B*27$^+$ patients with AAU, including eye-specific clonal expansion in a patient without arthritic manifestations of AS.

As the peptides we identified so far are not known to be linked to AS or AAU, the question remains whether potential disease-triggering peptides are within the list reported here. They are not obvious in peptide elution studies of HLA-B*27-transfected cell lines or of spleen cells from HLA-B*27 transgenic rats[42–45]; however, low levels of B*27-associated self-peptides, undetectable by mass spectrometry, could be antigenic to T cells previously primed by cross-reacting bacterial peptides. GPER1 is widely expressed in human tissues, including cartilage[46], bone[47] and the eye[48]. Within the eye, its expression is largely restricted to the iris and ciliary body, which is inflamed during AAU[48]. The inner-membrane YEIH protein with putative sulfate exporter function is conserved across several Gram-negative bacteria, of which many are known or suspected ‘triggers’ for ReA.

We solved seven crystal complex structures to visualize TCR and pMHC interactions. All TCRs use similar solutions to interact with the conserved peptide hotspots. This structural motif could extend beyond the peptides identified here through molecular mimicry. These findings broaden the concept of the arthritogenic peptide, suggesting that there may be multiple microbial triggers with shared structural features. Likewise, several self-peptides might be involved in cross-reactivity, with some antigens derived from the iris or ciliary body in the eye and others from entheses or joints.

Our data highlight a core motif for potential arthritogenic peptides targeted by AS-enriched BV9$^+$ T cells and are in agreement with previous studies suggesting environmental pathogens as putative triggers for autoimmune diseases[49,50]. These findings add an important piece to the complex, 50-year-old puzzle of why HLA-B*27 is strongly associated with AS. An initial T cell-mediated event could precipitate an inflammatory response that becomes independent of the early antigenic trigger. Alternatively, such T cells might function as a disease amplifier within a pre-existing inflammatory process. For the former, longitudinal studies of HLA-B*27$^+$ individuals tracking these T cells from disease onset to remission with flares would present substantial challenges, but with great clinical and scientific rewards.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-022-05501-7.

Methods

Patients

Patients with HLA-B*27-associated AAU met the Standardization of Uveitis Nomenclature Working Group classification criteria[51]. All patients were HLA-B*27:05$^+$ by HLA sequencing, performed by Histogenetics LLC (Ossining, NY). Two patients (UV027 and UV122) also met the Assessment of SpondyloArthritis International Society classification criteria for axial spondyloarthritis, with UV027 having radiographic disease (that is, AS)[52], whereas UV019 and UV180 did not have axial symptoms or radiographic findings. Each sample from the same patient is denoted by an '_a' or '_b' suffix to indicate whether it was the first or second collection, respectively. Aqueous humour and blood samples were collected during active anterior uveitis defined by at least 1+ anterior chamber (AC) cell (>6 cells per high-powered field)[53]. AC sampling was carried out as previously described[54]. In brief, about 100–200 µl of AC fluid was removed and centrifuged at 400*g* for 5 min, after which the aqueous fluid was removed. Blood samples were obtained by venipuncture, collected into EDTA tubes, and purified by Ficoll-Hypaque density gradient centrifugation. The AC cells and PBMCs were cryopreserved in FBS with 10% DMSO and stored at −140 °C.

HLA-B*27:05$^+$ patients with AS fulfilling modified New York criteria and attending the Oxford University Hospitals NHS Foundation Trust were recruited with ethically approved consent and protocols (ethics reference number 06/Q1606/139). A second cohort, approved by the local ethics committee of the Russian National Research Medical University in compliance with the Russian Ministry of Health, and processed in Oxford according to the Human Tissue Act 2004 handling and storage guidelines, was also included. Both synovial fluid mononuclear cells and PBMCs were fractionated by density centrifugation and cryopreserved before downstream flow sorting and single-cell sequence analyses.

Cell sorting of patient PBMC samples

Samples from HLA-B*27:05$^+$ patients with AS were subjected to cell sorting in two rounds. In the first round, three patient PBMC samples of 8–10 million cells were thawed, centrifuged at 400*g* for 10 min at room temperature, washed and resuspended in R10 (RPMI + 10% FCS) and then rested overnight in an incubator at 37 °C and 5% CO_2. Cells were washed with PBS and stained with anti-CD3–FITC, CD8–PE, BV9–PerCPCy5.5 and BV510 Aqua Live/Dead stain, filtered through a BD Falcon 35-µm filter cap and sorted using a BD FACSAria III (J. Webber, Cell Sorting Facility, Kennedy Institute of Rheumatology, University of Oxford). In this round, cells were enriched for, but not

exclusively, $BV9^+$ TCRβ chains, and approximately 4×10^5 cells per sample were submitted for library generation.

In a second round, four patient PBMC samples were thawed, washed and rested as above, and then stimulated for 28 days with 30 ng ml^{-1} anti-CD3 (Clone HIT3a, BD Pharmingen) and 50 U ml^{-1} recombinant IL-2 (Peprotech). Cells were fed fresh medium + IL-2 every 3–4 days and restimulated with anti-CD3 after 14 days. Following expansion, cells were subjected to magnetic depletion using microbeads coated with anti-CD19, anti-CD4 and anti-CD14 antibodies (BD Biosciences) according to the manufacturer's protocol (MACS Cell Separation – Miltenyi Microbeads and Columns, Miltenyi Biotec). Cells were rested overnight in R10 and then stained and sorted as above. In this round, cells were enriched exclusively for $BV9^+$ TCRβ chains and $0.5–5 \times 10^5$ cells per sample were submitted for library generation.

TCRα and TCRβ repertoire sequencing

Parallel TCRα and TCRβ chain cDNA libraries from the same cell sample were prepared using a previously described protocol[55]. The resulting libraries were sequenced on the Illumina NextSeq with 2 × 150-bp sequencing length. MiGEC[56] and MiXCR[57] software programs were used for sequencing data preprocessing, unique molecular barcode (UMI)-based normalization, and extraction of TCR sequences and quantitative information. Only those TCR cDNA sequences that were supported by at least two independent sequencing reads were used.

Single-cell TCR sequencing for AS cells

scRNA-seq data were generated using the 10x Chromium platform (10x Genomics). Libraries were generated from sorted cell samples by the Oxford Genomics Centre, Wellcome Centre for Human Genetics, sequenced by HiSeq4000 150-bp PE and subjected to chromium VDJ solution using the 10x Cell Ranger pipeline. Preliminary data analysis was conducted using Loupe V(D)J Browser and Loupe Cell Browser software (10x Genomics).

Single-cell TCR sequencing for AC cells

Frozen AC cells were thawed and washed once with FBS and once with PBS with 0.1% BSA. Owing to low cell numbers (5,000–30,000 cells), minimal processing was carried out to reduce cell loss. PBMCs were thawed and washed twice with 10% RPMI and then once with PBS with 0.1% BSA. Viability was >95% by trypan blue exclusion. Single-cell 5′ gene expression cDNA libraries were generated using the Chromium Controller (10× Genomics) platform for microdroplet-based, single-cell barcoding, at the Genome Technology Access Center at the McDonnell Genome Institute (GTAC@MGI, Washington University in St Louis). T cell enrichment libraries were generated with the Chromium Single Cell V(D)J Enrichment Kit (10× Genomics) at GTAC@MGI. Libraries were sequenced on the NovaSeq Sequencing System (Illumina). Sequencing reads were aligned to the human genome (GRCh38) using with Cell Ranger VDJ command (v4.0.0, 10× Genomics). Only clonotypes with one or two productive rearrangements for TCRα and one productive rearrangement for

TCRβ were used. Clonotypes were matched across tissues on the basis of on usage of TRAV and TRBV gene segments and CDR3α and CDR3β amino acid sequences.

Evolving HLA-B*27:05 variant for yeast display

The SRYWAIRTR peptide, β_2m and the HLA-B*27:05 heavy chain were expressed as a linker-interspersed single-chain trimer (SCT) that included a linker-accommodating HLA-B*27:05(Y84A) alteration[33,35,58,59]. This SCT was expressed at a low level on the yeast surface, indicated by a lack of both control GRb TCR (specific for the SRYWAIRTR peptide) tetramer and anti-MYC tag antibody staining (Extended Data Fig. 3a). As this probably reflected inefficient protein folding, alterations were introduced into the entire SCT reading frame by error-prone PCR (Extended Data Fig. 3b). After four rounds of selection, the MYC epitope and GRb TCR staining were detected, indicating improved SCT expression (Extended Data Fig. 3c). Sequence analysis revealed HLA-B*27:05 heavy chain alterations, specifically M5L, H114Y, A153D and R169S. As the MHC α2 helix R169S substitution represents a potential TCR contact residue, S169R reverse mutagenesis was carried out, and this retained GRb TCR tetramer staining (Extended Data Fig. 3c). The final SCT construct included L5M, H114Y and A153D substitutions for optimal expression, referred to herein as the HLA-B*27:05$^{(3mut)}$ display construct (Extended Data Fig. 3d).

Yeast display library preparation and selection

Yeast display libraries were prepared as previously reported[60]. In short, primers encoding chosen codons were used to make peptide libraries. To allow optimized HLA-B*27:05 binding, peptide anchor positions at P2 and PΩ were limited to Arg and Arg-Lys whereas other positions were encoded with NNK codons to allow all possible amino acids. Two peptide library lengths (9-AA and 10-AA; Extended Data Fig. 3e) were constructed separately and fused with two different tags (MYC and HA) for detection by flow cytometry. The diversity of the yeast libraries was determined by colony counting after electroporation.

Transformed yeast libraries were grown in SDCAA pH 4.5 and frozen down in aliquots at 10× diversity (that is, 10^{10} cells per aliquot). The yeast libraries were induced for protein production in SGCAA pH 4.5 and selected as previously described. In brief, the 10× diversity yeast libraries were first incubated with streptavidin-coated magnetic MACS beads (Miltenyi) for background clearance. After 1 h of incubation at 4 °C, yeast cells were passed through an LS column (Miltenyi) on a magnetic stand (Miltenyi) and washed three times. The flow-through was collected and then incubated with 400 nM biotinylated soluble TCR coupled to streptavidin-coated magnetic MACS beads for 3 h at 4 °C. Yeast cells were passed through a new LS column on a magnetic stand and washed three times. The flow-through was discarded, and the yeasts bound on the LS column were eluted and grown in SDCAA at pH 4.5 overnight. Confluent yeast culture was then induced in SGCAA pH 4.5 for subsequent selections. After four rounds of selection, all rounds of yeast were stored at 4 °C for deep-sequencing preparation.

Deep sequencing of pHLA libraries

DNA from 5×10^7 yeast cells from each round of selection was isolated by miniprep (Zymoprep II kit, Zymo Research). The flanking region of the sequencing product contained

random 8-nt sequences and individual barcodes to separate different rounds of yeast selection products. Sequencing products containing the yeast signal peptide, peptide libraries and N-terminal section of β_2m were PCR amplified for 25 cycles. The libraries were purified by agarose gel purification and quantified by nanodrop. Deep sequencing was carried out by an Illumina Miseq sequencer using a 2× 150 V2 kit. Paired-end reads were identified and processed for further analysis from deep sequencing using PandaSeq. Paired-end reads were separated according to designated barcodes using Geneious V6. Reads with frameshifts or stop codons were removed from further analysis. Identical peptide sequences were quantified and clustered using Perl scripts and shell commands[60]. To generate position frequency matrices for each TCR, we restricted MHC anchor residues to Arg, Gln and Lys at P2 and Arg, Lys, Met, Ile, Leu, Tyr, Phe, Trp and Val at P9. Each position frequency matrix was used to search for peptides from the human proteome and from known microbial ReA triggers (*Chlamydia*, *Klebsiella*, *Salmonella*, *Shigella* and *Yersinia*) in the UniProt database[4,29–31].

T cell activation in co-culture assays

Eleven AS (blood and synovial) and two ocular-derived TCRs were cloned into a lenti-viral based vector pHR. The full-length sequences of the α- and β-chains were cloned separately and transduced with pMD2G and pSPax packaging vectors. Lentivirus was generated for each TCR in embryonic kidney 293T cells using Fugene in complete DMEM containing 10% FBS. Virus was collected at 72 h each in 2.5–4 ml medium and pelleted to remove cell debris before infection of 1 million $CD8^+$ SKW-3 cells. After 5–7 days of expansion in RPMI complete medium containing 10% FBS, cells were stained for CD3 (UCHT1, BioLegend) or TCR (IP26, BioLegend) and sorted on the SH800 cell sorter (Sony Biotechnology).

The full-length sequences of canonical HLA-B*27:05 and β_2m were separately cloned into pHR vectors and transduced with pMD2G and pSPax packaging vectors. Lentivirus was generated for HLA-B*27:05 in embryonic kidney 293T cells using Fugene in complete DMEM containing 10% FBS. Virus was collected at 72 h in 2.5–4 ml of medium and pelleted to remove cell debris before infection of 1 million K562 cells. After 5–7 days of expansion in RPMI complete medium containing 10% FBS, cells were stained for HLA expression (anti-human HLA-A, B, C antibody, clone W6/32 BioLegend) and sorted on the SH800 cell sorter (Sony Biotechnology).

The transduced K562 cells expressing functional HLA-B*27:05 molecules were cultured with the top five peptides from each AS TCR selected yeast library or human and microbial endogenous peptides for 1–2 h at 37 °C with 5% of CO_2. The peptides were synthesized to crude levels (Elim Biopharm) and were dissolved in DMSO at 10 mM concentration. A 1 μl volume of dissolved peptides was added into 100 μl of cells to reach a 100 μM final concentration. After an incubation period, cells were washed with RPMI complete medium once and then incubated with SKW-3 T cells for 14–18 h at 1:1 ratio. CD69 (FN50, BioLegend) and CD3 (OKT3 or UCHT1, BioLegend) expression was monitored by flow cytometry using CytoFlex. Peptide dose–response experiments were carried out with peptides synthesized to 80% purity (Genscript) and carried out in triplicate (biological

and technical). For the antigen-processing experiment, HLA-B*27:05-K562-expressing cells were further transduced with 60 flanking amino acids that encompassed the GPER1 or YEIH epitope. K562 cells that co-express HLA-B*27:05 and GPER1 or YEIH were subsequently co-incubated with AS8.4 TCR SKW-3 T cells. CD69 (FN50 BioLegend) and CD3 (OKT3 or UCHT1 BioLegend) expression was monitored by flow cytometry using CytoFlex software CyteExpert. *P* values were calculated by ordinary one-way analysis of variance in Prism with standard deviations.

Expression of AS TCR as inclusion bodies in *E. coli*

AS2.1, AS3.1 or AS4.1 TCR α- and β-chains were cloned into the pET22b expression vector through NdeI and XhoI restriction sites, incorporating a stop codon for the α-chain and a BirA tag and stop codon for the β-chain, respectively. A second beta chain construct omitted the BirA tag for production of untagged protein for surface plasmon resonance (SPR) analysis.

TCRα or TCRβ–pET22b constructs were used to transform the *E. coli* BL21 (DE3) pLysS-derived Rosetta strain, under chloramphenicol and ampicillin selection, for protein expression in inclusion bodies[61]. In brief, 40 ml of overnight starter culture in LB medium with chloramphenicol and ampicillin was divided between 4 l of low-salt LB (5 g NaCl per litre) in four 2-l Erlenmeyer flasks and grown for about 2–3 h to an optical density at 600 nm of 0.4–0.6, before induction with 0.5 mM IPTG. Cells were collected after 4 h expression by centrifugation at 6,000*g* for 15 min, resuspended in about 50 ml final volume of PBS and stored frozen at −20 °C at least overnight.

Following thawing, cells were transferred to a glass beaker within an ice box and subjected to sonication using an ultrasonicator (Cole-Palmer) with a flat head probe and pulse settings of 30 s on/30 s off at an amplitude of 30% for a total sonication on time of 5 min. Cellular debris and proteins were transferred to pre-chilled Oakridge tubes and pelleted by centrifugation at 15,000*g* for 15 min at 4 °C, the supernatant was discarded and the pellet was washed by resuspension in ice-cold detergent wash buffer (50 mM Tris-HCl pH 8, 200 mM NaCl, 10 mM EDTA pH 8, 0.5% Triton X-100, 0.01% NaN_3, fresh dithiothreitol to 2 mM) using a Dounce glass homogenizer. Centrifugation and resuspension were conducted an further two times to wash away host proteins and impurities before a final wash step using resuspension wash buffer (the same as detergent wash but without Triton X-100).

After final centrifugation, the pellet was dissolved in about 20 ml of urea solubilization buffer (8 M urea, 50 mM MES pH 6.5, 10 mM EDTA pH 8, fresh 2 mM dithiothreitol) by rotation at 4 °C overnight. Following centrifugation at 15,000*g* for 30 min the next day to remove any final insoluble debris, inclusion bodies were quantified by measuring absorbance at 280 nm, aliquoted and stored at −80 °C.

Chemical refolding of AS TCR from inclusion bodies and chromatographic purification

Chemical refolding was conducted essentially as previously published[61], but with minor modifications. A 25 mg quantity of each inclusion body's purified TCR α- and β-chains was diluted in 500 ml of denaturing refolding mix (5 M urea, 0.4 M L-arginine, 100 mM Tris-HCl, 2 mM EDTA, 5 mM reduced glutathione, 0.5 mM oxidized glutathione)

with stirring overnight at 4 °C. The refolding mix was then transferred to dialysis tubing (Spectrapor RC 10 kDa MW cutoff) and dialysed against 5 l of distilled H_2O with stirring at 4 °C for 3 days, followed by a final 24-h dialysis against 5 l of 10 mM NaCl, 10 mM Tris-HCl pH 8 to buffer exchange in preparation for ion-exchange chromatography.

Ion-exchange chromatography was conducted using a HiTrap Q FF 5-ml anion-exchange column attached to an Akta Start chromatography system (GE Healthcare), equilibrated with 10 mM NaCl, 10 mM Tris pH 8. After sample loading and a 10-column-volume (CV) wash step, protein was eluted using a gradient of 30 CV to 50% 1 M NaCl, 10 mM Tris pH 8. SDS–polyacrylamide gel electrophoresis (PAGE) under reducing and non-reducing conditions was used to analyse chromatography peaks and verify correct heterodimer formation.

Fractions containing αβ TCR heterodimer were pooled and concentrated using centrifugal concentrators (10-kDa MW cutoff) to a volume of about 6 ml, and then subjected to buffer exchange using PD10 columns into 10 mM Tris pH 8 to remove salt, before enzymatic biotinylation overnight at room temperature through BirA-500: BirA biotin-protein ligase standard reaction kit (Avidity LLC).

As a final purification step to remove aggregates and excess biotin, protein was then subjected to gel filtration chromatography against 150 mM NaCl 20 mM Tris pH 8 using a HiLoad 26/600 Superdex 200 prep grade column (GE Healthcare). Following SDS–PAGE under reducing and non-reducing conditions to identify fractions containing αβ TCR, these fractions were pooled and concentrated, and quantified by measuring absorbance at 280 nm, and biotinylation efficiency was estimated by addition of streptavidin followed by a second round of SDS–PAGE. Protein was then aliquoted and stored at −80 °C.

Expression of AS TCR in Expi293 mammalian expression system

AS TCR α-chains were cloned into the pD649 expression vector encoding a C-terminal basic helix zipper, and β-chains were cloned into the same vector encoding a C-terminal acidic helix zipper and a BirA biotinylation tag. Transfection and expression in the Expi293 system was conducted as per the kit instructions (Expi293 Expression System, ThermoFisher Scientific). For in situ biotinylated proteins, 25 ml of Expi293F cell culture at a density of 3×10^6 cells ml^{-1} was transfected with 13.5 μg of each TCR chain, plus 3 μg of pHLSec vector[62] encoding BirA with a KDEL ER-retention signal, and expression proceeded for 4 days.

Following clarification of supernatant by sequential centrifugation at 400*g* for 10 min at 4 °C to remove cells, and then 3,000*g* for 20 min at 4 °C to remove debris, AS TCR heterodimers were purified by Ni-NTA affinity chromatography using a HisTrap Excel 1-ml column attached to an Akta Pure chromatography system (GE Healthcare). Supernatant was diluted approximately 1.5× in binding buffer (20 mM NaPi pH 7.4, 0.5 M NaCl) to ensure efficient binding and loaded onto the column, followed by 5 CV of wash with binding buffer and elution in 5 CV of elution buffer (20 mM NaPi pH 7.4, 0.5 M NaCl, 0.5 M imidazole). The elution peak was concentrated by a centrifugal concentrator to a volume of 1 ml and loaded onto an S200 HR 10/300 gel filtration column (GE Healthcare)

equilibrated with 150 mM NaCl 20 mM Tris pH 8. Fractions were analysed for the presence of heterodimeric TCR and biotinylation by streptavidin shift assay and SDS–PAGE under reducing and non-reducing conditions. Pooled fractions were concentrated, quantified by measuring absorbance at 280 nm, and aliquoted and stored at −80 °C.

AS TCR tetramer staining of pHLA-B*27 SCT-transfected 293T Δ*B2M* cells

TCR tetramers were generated by incubation at a final molar ratio of 4:1 with streptavidin–APC (Invitrogen), added gradually over 2 h in six pulses. Tetramers were used to stain Δ*B2M* 293T cells transfected with SCT constructs encoding peptide, β_2m and HLA-B*27:05 heavy chain fused to eGFP in the pGEN plasmid backbone.

In brief, Δ*B2M* 293T cells were grown in 6-well dishes to approximately 60–70% confluency and transiently transfected using Genejuice reagent (Merck Millipore). Cells were assessed for eGFP expression by confocal microscopy at 24 h post-transfection, collected, washed once in ice-cold PBS, and then stained with 1 μg of tetramerized TCR per well for 30 min at 4 °C. For assessment of pMHC expression, cells were separately stained with anti-HLA-B*27 antibody clone ME1 (in-house hybridoma[63]) for 15 min at 4 °C, washed and then stained with secondary goat anti-mouse Fc–APC; or concomitantly with directly conjugated anti-β_2m monoclonal antibody 2M2–PEcy7 (BioLegend) for 15 min at 4 °C. Cells were then washed and resuspended in PBS for flow cytometry using a CyAn ADP flow cytometer (Beckman Coulter). Analysis was conducted using FlowJo 10 (Becton Dickinson).

Expression and purification of HLA-B*27 complexes

HLA-B*27 complexes with AS-related peptides were generated essentially as previously described[64], with some modifications. HLA-B*27:05, HLA-B*27:06, HLA-B*27:09 and HLA-B*27:05(H114Y) heavy chains and β_2m were expressed from the pET22b expression vector and inclusion bodies were purified as described above. Refolding was by staged dilution into 200 ml refolding mix (0.4 M l-Arg, 100 mM Tris pH 8, 2 mM EDTA, 5 mM reduced glutathione, 0.5 mM oxidized glutathione) at 4 °C with stirring. First, 2 mg of β_2m was refolded for 30 min, followed by the addition of 15 mg DMSO-solubilized peptide and subsequently 2 pulses of 5 mg HLA-B*27 heavy chain over a period of 48 h. The refolding mix was then vacuum filtered through a 0.22 μM Durapore PVDF membrane (Merck Millipore) and concentrated using a Vivaflow 50R cross-flow filtration system to a volume of about 30 ml followed by centrifugal concentration to 3 ml. Buffer exchange by PD10 and biotinylation was carried out as described above. Refolded pMHC complexes were then purified by gel filtration on a Superdex 75 16/60 column equilibrated with 150 mM NaCl 20 mM Tris pH 8. HLA-B*27 complexes used for crystallographic, SPR and thermal melt studies incorporated a C-to-S modification at position 67 on the HLA-B*27 heavy chain to reduce heavy chain dimer formation and to improve refolding yield.

Thermal melt analysis of HLA-B*27 subtypes in complex with bacterial- and self-derived peptide

The thermal stabilities of HLA-B*27:05, HLA-B*27:05(H114Y), HLA-B*27:06 and HLA-B*27:09 refolded with the library-mined PRPF3^{self}, GPER1^{self} and YEHI^{bac} were

determined by differential scanning fluorimetry using a Prometheus Panta instrument (Nanotemper). A 0.50 μg μl^{-1} concentration of refolded HLA-B*27 material was diluted in 20 mM Tris pH 8, 150 mM NaCl to a final volume of a 20 μl. Individual samples were split between two Prometheus Grade Standard Capillaries (Nanotemper) and transferred into a capillary sample holder. Excitation power was pre-adjusted to obtain a range between 8,000 and 15,000 raw fluorescence units for fluorescence emission detection at 330 nm and 350 nm. A thermal ramp of 1 °C min^{-1} from 20 °C to 95 °C was applied. Automated thermal melt data calling was achieved using the Thermal Unfolding Analysis software within PR.Panta Analysis, version v1.2.

SPR analysis of AS TCR-pHLA-B*27:05 interactions

SPR analysis was conducted using a BIAcore S200 system (GE Healthcare), using HBS-EP^{+} (10 mM HEPES, 150 mM NaCl, 3 mM EDTA, 0.05% v/v surfactant P20) as a running buffer. Biotinylated peptide–HLA-B*27:05 complexes were immobilized onto three flow cells of a CM5 Series S SA (streptavidin-coated) sensor chip, or on a reusable Series S Sensor Chip CAP (according to the instructions in the Biotin CAPture kit, GE Healthcare). Between 500 and 2,000 RUs were immobilized in separate experiments. An irrelevant protein was immobilized at similar levels in one flow cell to act as a control surface. Equilibrium binding analyses of AS TCR–pHLA-B*27:05 interactions were carried out at 20 °C by measuring binding responses at equilibrium of serial injections (twofold dilutions) of soluble non-biotinylated AS TCRs, from low to high concentration, at a flow rate of 30 μl min^{-1}. GraphPad Prism 8 software was used to conduct curve fitting and derive the equilibrium dissociation constant, K_D.

Crystallization of AS TCR–pHLA–B*27:05 complexes

AS TCRs produced from GNTI^{-} Expi cells were treated with Endo H, 3C protease, and carboxypeptidase A and B to remove sugars and C-terminal zippers for 2 days at 4 °C and subsequently purified on an S200 column. Purified AS TCRs and pHLA-B*27:05 were mixed at 1:1 ratio at 10 mg ml^{-1} concentration for crystallization screening. AS3.1–PRPF3^{self}–B*27:05 complex was grown in 25% PEG3350 with 0.2 M ammonium tartrate dibasic dihydrate; AS4.2–PRPF3^{self}–B*27:05 complex was grown in 20% PEG3350 with 0.2 M potassium phosphate monobasic; AS4.3–PRPF3^{self}–B*27:05 complex was grown in 25% PEG3350 with 0.2 M ammonium sulfate and 0.1 M Bis-Tris pH 6.5; AS4.2–YEIH^{bac}–B*27:05 complex was grown in 20% PEG3350 with 0.2 M ammonium acetate; AS4.3–YEIH^{bac}–B*27:05 complex was grown in 25% PEG3350 with 0.2 M ammonium sulfate and 0.1 M Bis-Tris pH 6.5; AS4.3–RNASEH2B^{self}–B*27:05 complex was grown in 20% PEG3350 with 0.2 M magnesium formate dihydrate; AS8.4–YEIH^{bac}–B*27:05 complex was grown in 25% PEG3350 and 0.1 M HEPES pH 7.5. All crystals were snap frozen in mother liquor containing 30% glycerol. Diffraction data were collected at Stanford Synchrotron Radiation Laboratory beamlines 12–1 and 12–2. Datasets were indexed and scaled with XDS.

Structure determination and refinement

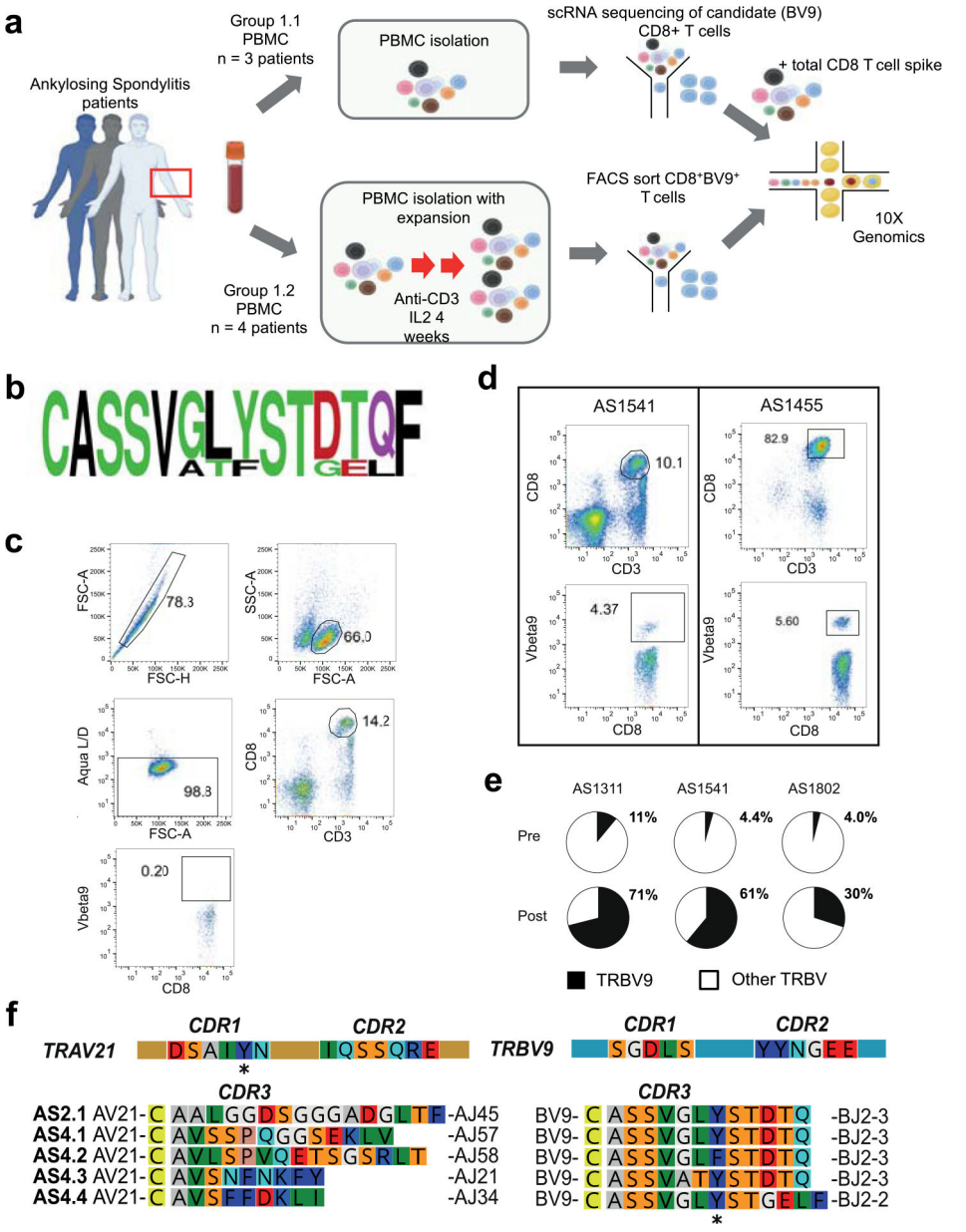
All structures were solved through molecular replacement with the PHASER program. The molecular search model for HLA-B*27:05 is 2BST and the molecular search model for TCR

is 5KSB (refs. [32,36]). The MHC molecule and TCR CDR3 loops are removed in 5KSB before the search to reduce model bias. The refined model is then used as a search model for the rest of the AS TCR-pHLA-B*27:05 complex structures. All structures were refined with the PHENIX software suite and manually modified with the COOT program[65,66]. Structural analysis was carried out under programs within the CCP4 package. The statistics for data collection and refinement are summarized in Supplementary Table 6. Structures have been deposited in the RCSB protein data bank with the accession codes 7N2N, 7N2O, 7N2P, 7N2Q, 7N2R, 7N2S and 8CX4.

Statistics

Sample sizes were not pre-determined using statistical methods. Investigators were not blinded to the experimental sample identities. Peptide titration experiments were carried out, and antigen-processing experiments were carried out with biological triplicates. The pHLA-B*27 thermal stability assay was carried out with technical triplicates. *P* values are indicated in the figures as: NS, $P < 0.1$; *$P < 0.05$: **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$.

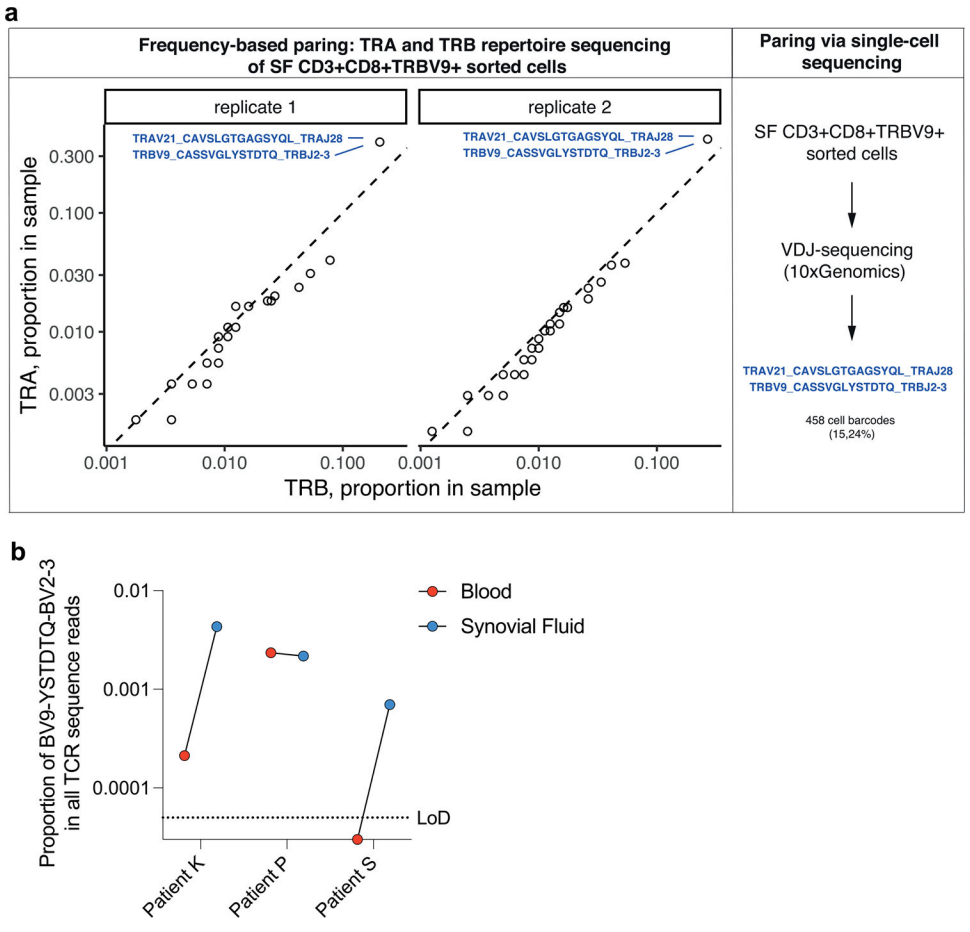
Extended Data



Extended Data Fig. 1 |. Isolation of TRBV9+ TCRs from PBMC.

a. Schematic view of TRBV9+ TCRs isolation strategy from patients' PBMC. Created with BioRender.com. b. Weblogo of CDR3β variants from HLA-B*27:05+ AS patient PBMC samples. c. Sorting strategy for scRNAseq library generation for Round 1 samples (HLA-B*27:05+ AS patient PBMC with overnight rest, post thaw). Sample AS1541 is shown. Final panel shows BV9-PerCP-Cy5.5 isotype control staining against CD8-PE. d. Representative sorting data for scRNAseq library generation. Sample AS1541 from Round 1 (HLA-B*27:05+ AS patient PBMC with overnight rest, post thaw) and sample AS1455B2 from Round 2 (HLA-B*27:05+ AS patient PBMC, 28 days expansion with αCD3 and IL-2 followed by magnetic depletion of CD4+, CD19+ and CD14+ cells) are shown. CD8+

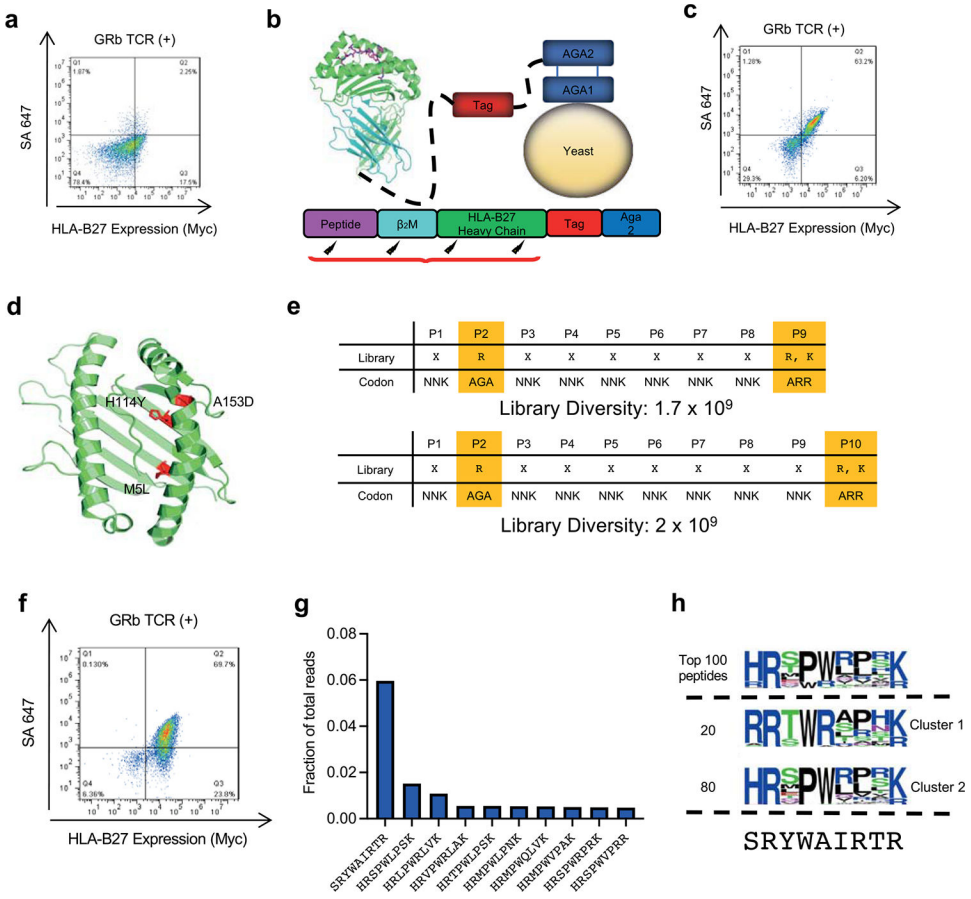
T cell frequencies (as a % of total live cells) and BV9+ frequencies (as a % of CD8+ T cells) for each sample are as follows: Round 1 samples (CD8+ %/BV9+ %): AS1802 14.9/3.97; AS1311 16.4/10.7; AS1541 10.1/4.37. Round 2 samples: AS1455 82.9/5.60; AS1567 31.4/1.16; AS1661 66.7/7.21; AS1803 49.4/12.2. e. Enrichment of TRBV9 during first round of scRNAseq library generation by sorting. Pie charts show percentage of BV9+ cells in each sample pre- and post-mixing of sorted fractions. f. Partial amino acid sequence alignment of five AS patient PBMC-derived TCRs in this study, in RasMol colouring. The CDR1 and CDR2 sequences are shown for TRAV21 and TRBV9 in the top row. CDR3 amino acid sequences are shown in the bottom rows along with corresponding variable and joining gene usage. Structurally important bulky residues are marked with asterisks.



Extended Data Fig. 2 |. AS3.1 TCR alpha-beta pair identification.

a. Correlation of the top100 TCR α and β nucleotide sequence proportion in two independent samples of $CD3^{+}CD8^{+}TRBV9^{+}$ synovial fluid T cells of patient P. Confirmation of the frequency-based paired TCR α and β with single-cell VDJ-sequencing of an independent sample of $CD3^{+}CD8^{+}TRBV9^{+}$ synovial fluid T cells from the patient. b. Enrichment of the BV9–Y/FSTDTQ–BJ2–3 motif in the joint compared to the blood in AS. Bulk TCR sequencing was performed on $CD8^{+}$ T cell cDNA from the blood and synovial fluid. The graph displays the proportion of TCR sequence reads containing the BV9–Y/FSTDTQ–BJ2–3 motif in paired blood and synovial fluid of three Ankylosing Spondylitis

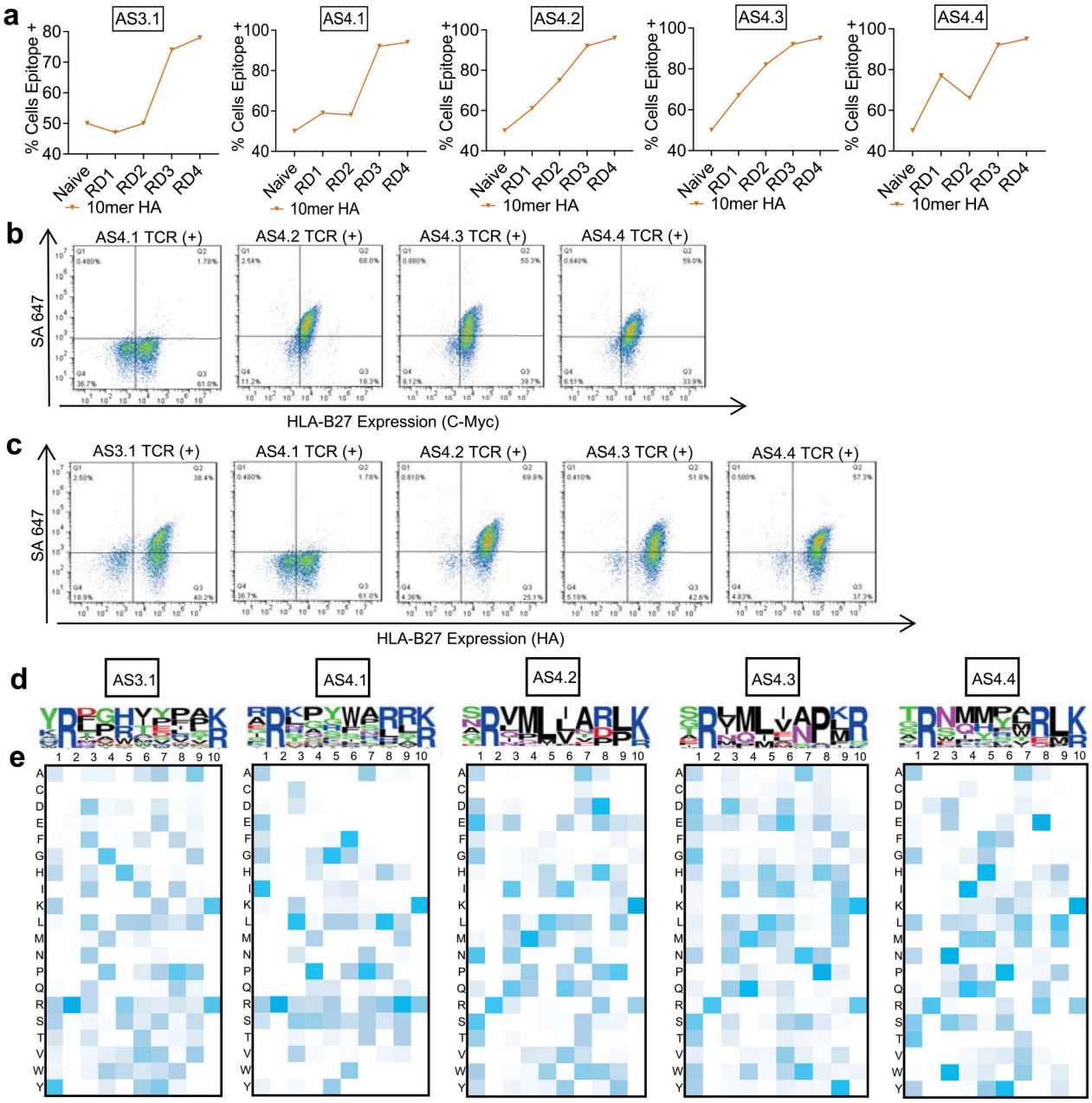
patients. Level of detection (LoD) of blood samples is indicated by a dashed line and is the median proportion of a single sequence read from all blood samples.



Extended Data Fig. 3 |. Development of HLA-B*27:05 yeast library.

a. GRb TCR tetramer staining of wildtype SRY-β_2M-HLA-B*27:05 single chain trimer expressed on yeast surface. Single chain trimer expression was monitored by anti-C-Myc-488. Streptavidin-647 (SA-647) was used to tetramerize and label the GRb TCR. b. Schematic of the SRY-β2m-HLA-B*27:05 single chain trimer on yeast surface. Error prone PCR introduced random mutations in SRY peptide, β2m and HLA-B*27:05 heavy chain. Sequencing of the enriched SRY-β2M-HLA-B*27:05 error prone library identified mutations located on HLA-B*27:05 heavy chain, shown as cartoon diagram and colored green. Bona fide mutations are shown as stick diagram and colored red. c. GRb TCR tetramer staining was rescued by M5L, H114Y and A153D mutation of the SRY-β2m-HLA-B*27:05 single chain trimer (HLA-B*27:05^{3mut}). Single chain trimer expression was monitored by anti-C-Myc-488. Streptavidin-647 (SA-647) was used to tetramerize and label the GRb TCR. d. Location of M5L, H114Y, and A153D mutations shown on the HLA-B*27 ribbon structure. e. Library design shows the anchor residue preference for the HLA-B*27:05^{3mut} library at the P2 and PΩ. For other positions, the NNK codon allowed all 20 amino acids and stop codons. Nucleotide abbreviations for codon usage are listed according to the IUPAC nucleotide code. Library diversities were estimated based on SDCAA plate colony numbers under serial dilution. f. GRb TCR tetramer staining of 4th

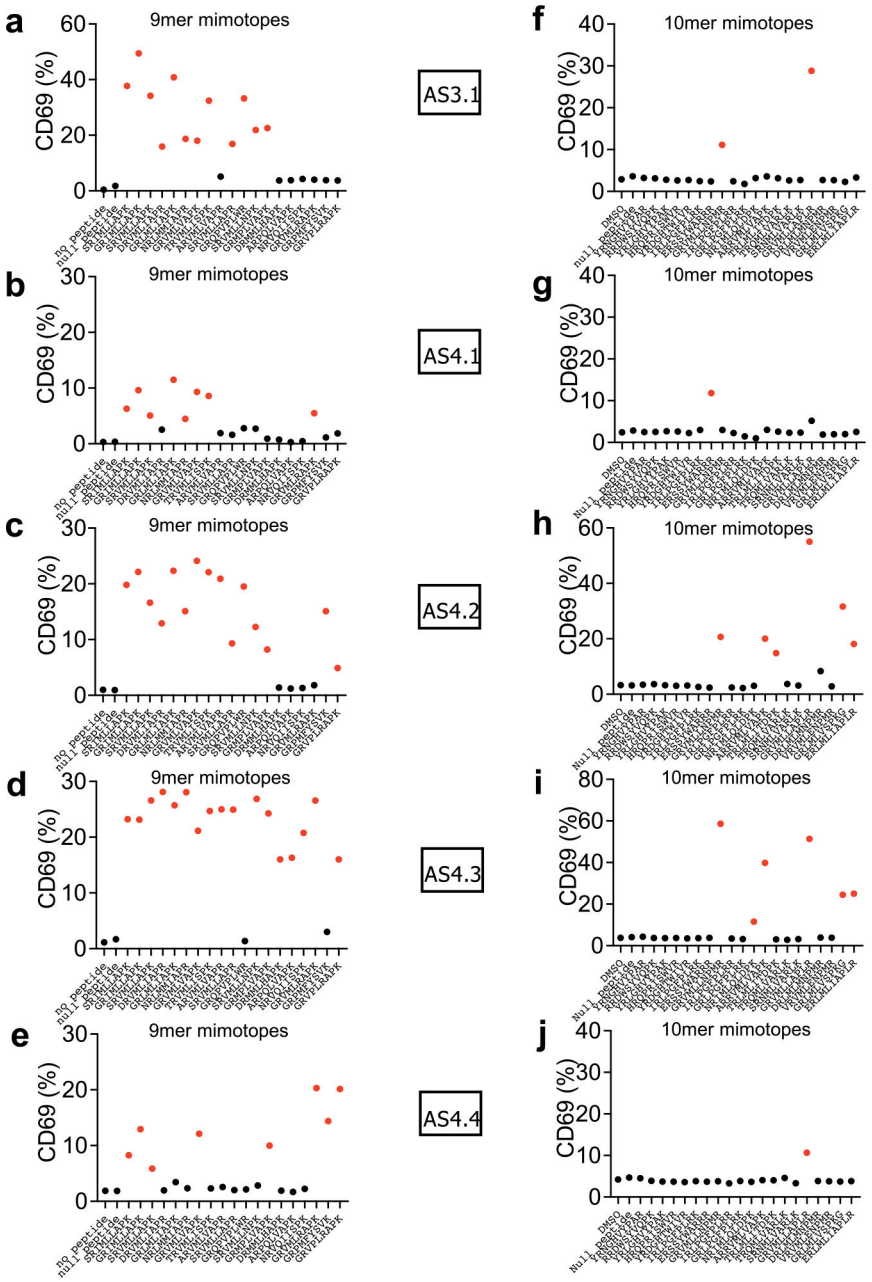
round peptide-HLA-B*27:05^{3mut} yeast-display library, indicating successful enrichment. g. Deep sequencing counts on the most enriched peptides after 4 rounds of selection with the GRb TCR. h. WebLogo plots of the top 100 peptides after 4th round of selection with the GRb TCR, the two predominant clusters within the selection, and the sequence of the cognate peptide recognized by the GRb TCR in the abundance at the given position among the unique peptides.



Extended Data Fig. 4 |. Selection of AS TCRs on HLA-B*27:05 libraries.

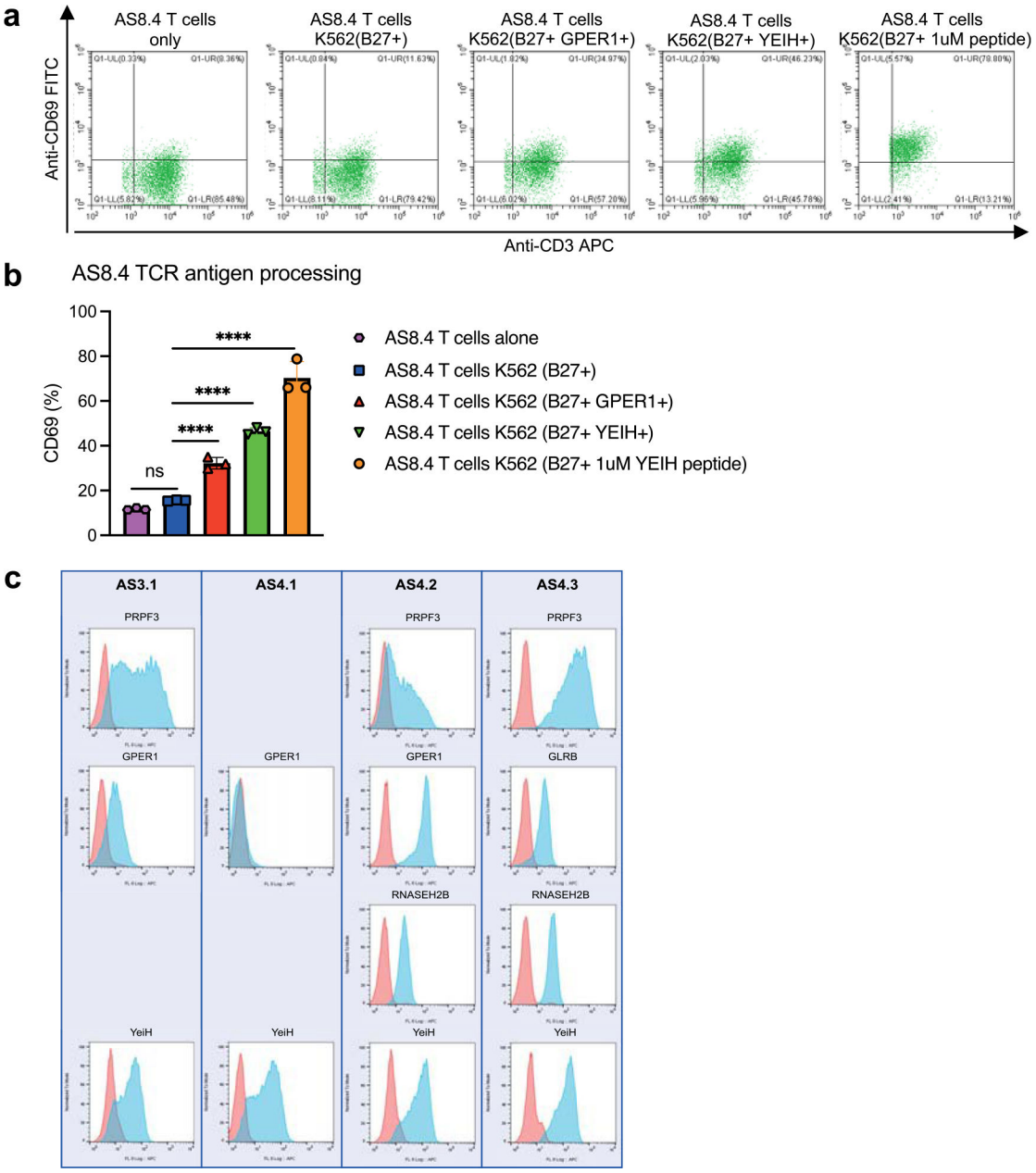
a. HLA-B*27:05^{3mut} 10mer library enrichment per round of selection by the five AS TCRs as measured by flow cytometry. b. AS4.1, AS4.2, AS4.3 and AS4.4 TCR tetramer staining on the 4th round HLA-B*27:05 9-AA libraries. HLA-B*27:05^{3mut} expression was

monitored by anti-C-Myc-488. Streptavidin-647 (SA-647) was used to tetramerize and fluorescently label the AS TCRs. c. AS3.1, AS4.1, AS4.2, AS4.3 and AS4.4 TCR tetramer staining on the 4^{th} round HLA-B*27:05^{3mut} 10-AA libraries. HLA-B*27:05 expression was monitored by anti-HA-488. Streptavidin-647 (SA-647) was used to tetramerize and fluorescently label the AS TCRs. d. WebLogos representing the unique 4^{th} round selected peptides for each AS TCR based on deep sequencing reads. The size of each amino acid letter represents its abundance at the given position among the unique peptides. e. Heatmap plots showing the amino acid composition per position of the peptides enriched after the 4^{th} round of selection. Darker color represents greater abundance of a given amino acid at specific position.



Extended Data Fig. 5 |. Synthetic peptide activation by AS TCR expressing T cell line.
(a-j) $CD8^+$ SKW-3 cells were transduced with AS TCRs via lentivirus and sorted for stable TCR (IP26) and CD3 (UTCH1) co-expression. K562 cells were transduced with wild type HLA-B*27:05 via lentivirus and sorted for stable HLA molecule expression. The antigen-presenting cell line was pulsed for 2 h with 100 μM peptides and co-incubated with the T cell lines for 18 h then analyzed for CD69 expression by flow cytometry. (a-e) AS3.1, AS4.1, AS4.2, AS4.3 and AS4.4 TCR were tested for CD69 activation by yeast-selected mimotopes from 9-AA libraries. (f-j) AS3.1, AS4.1, AS4.2, AS4.3 and AS4.4 TCR were tested for CD69 activation by yeast-selected mimotopes from 10-AA libraries. Red dots

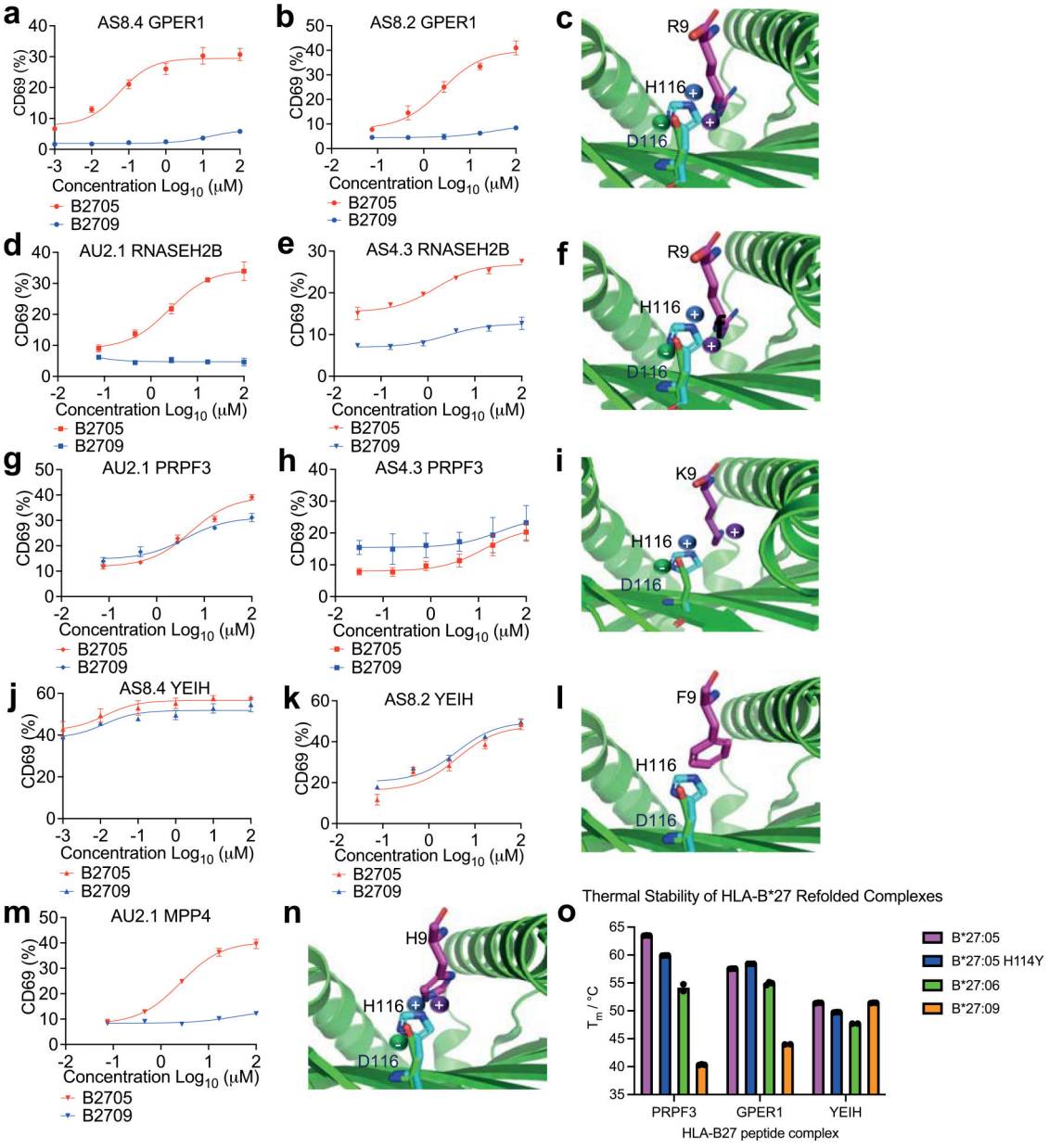
show activation, as defined by CD69% at least 2-fold greater than negative controls (DMSO and null peptide, n = 1).



Extended Data Fig. 6 |. AS TCR tetramer staining on SCT transfectants.

a. Representative flow cytometry showing AS8.4 cells alone, AS8.4 T cells + K562 (HLA-B*27:05^{+}), AS8.4 T cells + K562 (HLA-B*27:05^{+} GPER1^{+}), AS8.4 T cells + K562 (HLA-B*27:05^{+} YEIH^{+}) and AS8.4 T cells + K562 (HLA-B*27:05^{+} 1uM YEIH peptide). b. Plots of percentage of CD69% positive AS8.4 T cells alone, AS8.4 T cells + K562 (HLA-B*27:05^{+}), AS8.4 T cells + K562 (HLA-B*27:05^{+} GPER1^{+}), AS8.4 T cells + K562 (HLA-B*27:05^{+} YEIH^{+}) and AS8.4 T cells + K562 (HLA-B*27:05^{+} 1uM YEIH peptide). Data are shown as mean±s.d., n = 3 biological replicates. NS, not

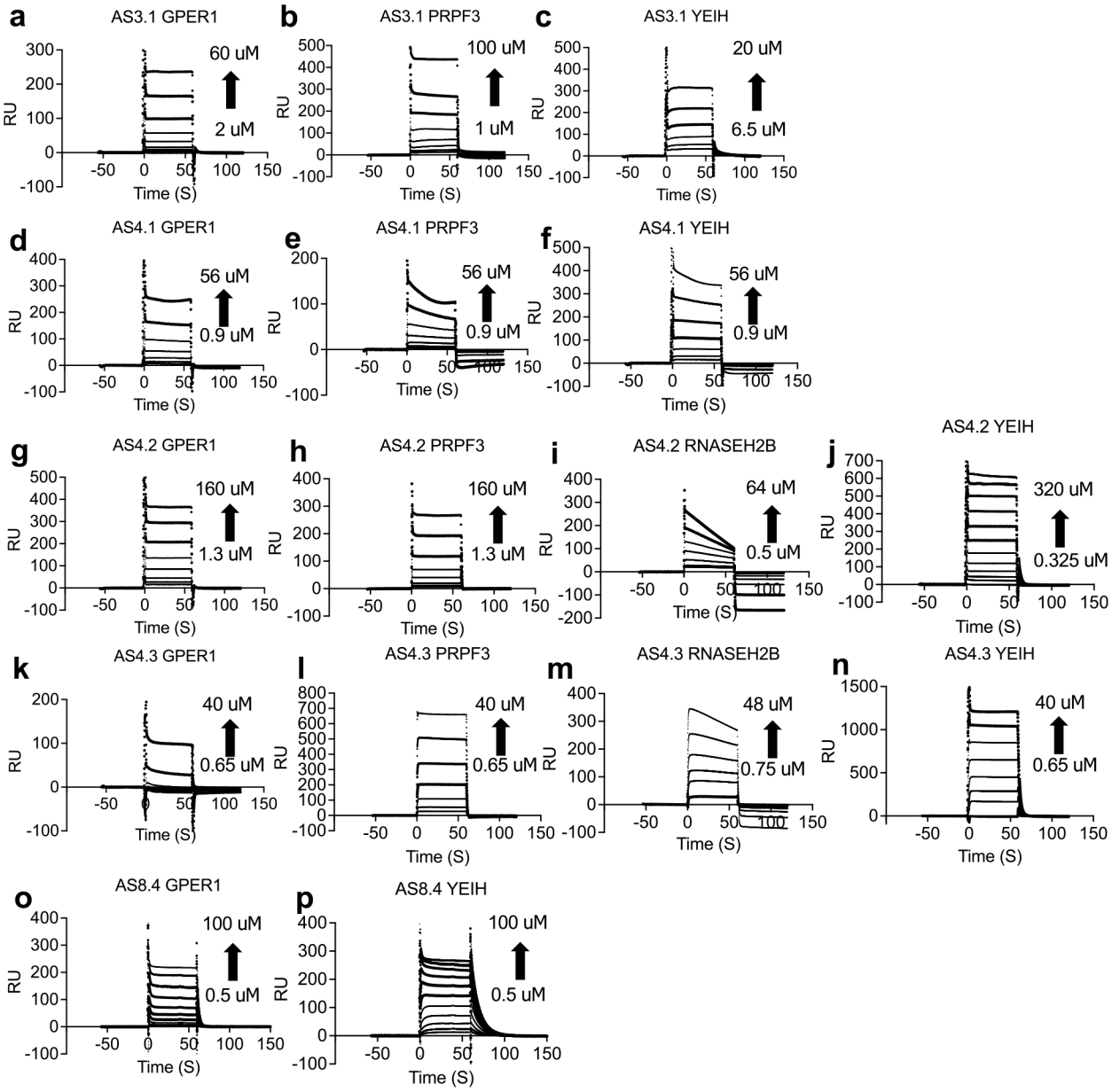
significant, P = 0.5453; ****P<0.0001 (one-way ANOVA). c. Representative flow cytometry histograms showing AS TCR tetramer staining for SCT transfected b2m KO 293T cells. Red histograms: negative controls (irrelevant p-HLA-B*27:05 construct); Blue histograms: potential arthritogenic peptide identified in this study.



Extended Data Fig. 7 |. Differential activation of AS TCRs by potential arthritogenic peptides on HLA-B*27:05 and HLA-B*27:09 backgrounds.

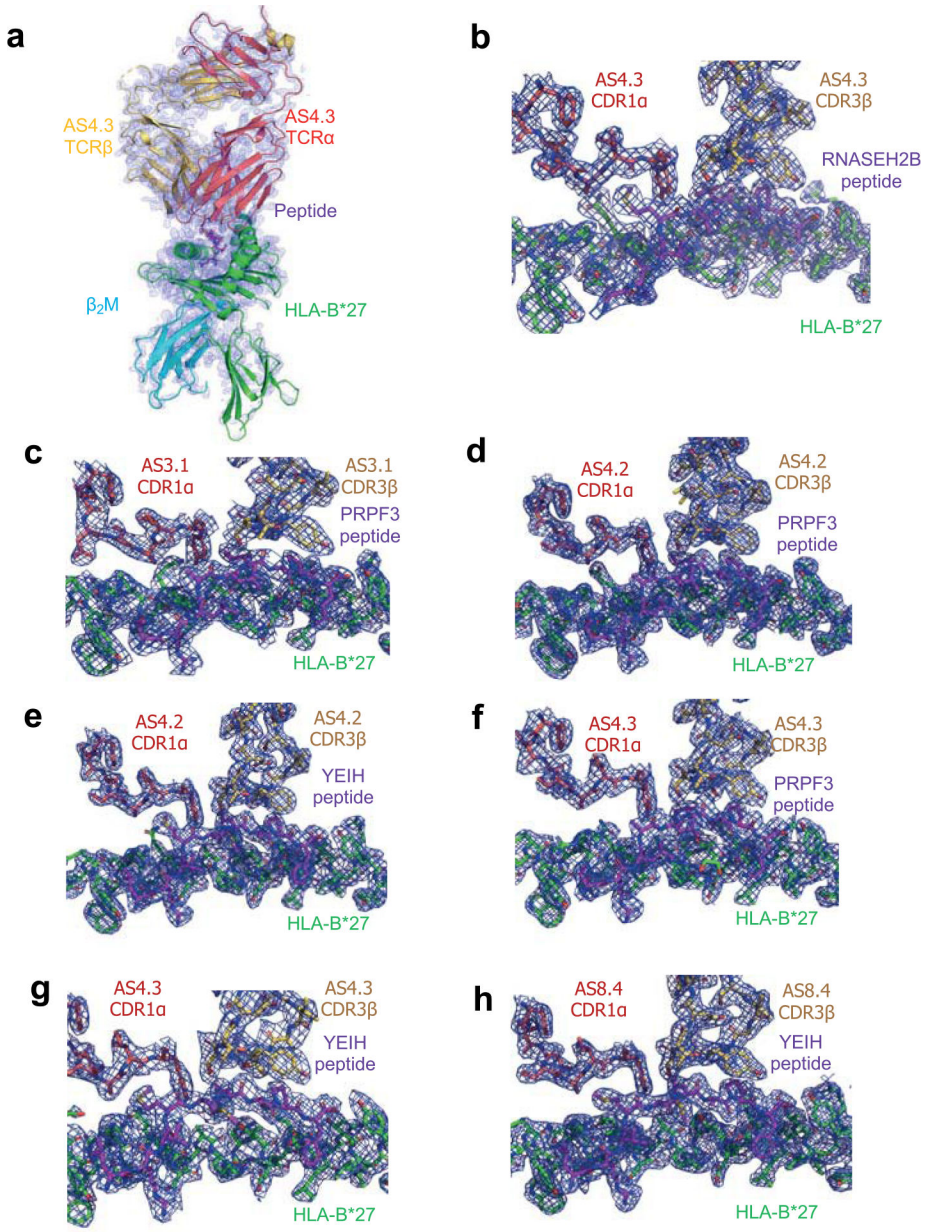
a. AS8.4 T cells incubated with GPER1-pulsed K562 cells, expressing HLA-B*27:05 or HLA-B*27:09, respectively, were tested for CD69 up-regulation. Data are shown as mean±s.d., n = 3 biological replicates. b. AS8.2 T cells incubated with GPER1-pulsed K562 cells, expressing HLA-B*27:05 or HLA-B*27:09, respectively, were tested for CD69 up-regulation. Data are shown as mean±s.d., n = 3 biological replicates. c. Structure

models showing how the GPER1 P9 Arg interacts with HLA-B*27:05 Asp at 116 and HLA-B*27:09 His at 116. The peptide Arg 9 is colored magenta, Asp 116 is colored green, and His 116 is colored cyan. d. AU2 T cells incubated with RNASEH2B-pulsed K562 cells, expressing HLA-B*27:05 or HLA-B*27:09, respectively, were tested for CD69 up-regulation. Data are shown as mean±s.d., n = 3 biological replicates. e. AS4.3 T cells incubated with RNASEH2B-pulsed K562 cells, expressing HLA-B*27:05 or HLA-B*27:09, respectively, were tested for CD69 up-regulation. Data are shown as mean±s.d., n = 3 biological replicates. f. Structure models showing how the RNASEH2B P9 Arg interacts with HLA-B*27:05 Asp at 116 and HLA-B*27:09 His at 116. The peptide Arg 9 is colored magenta, Asp 116 is colored green, and His 116 is colored cyan. g. AU2 T cells incubated with PRPF3-pulsed K562 cells, expressing HLA-B*27:05 or HLA-B*27:09, respectively, were tested for CD69 up-regulation. Data are shown as mean±s.d., n = 3 biological replicates. h. AS4.3 T cells incubated with PRPF3-pulsed K562 cells, expressing HLA-B*27:05 or HLA-B*27:09, respectively, were tested for CD69 up-regulation. Data are shown as mean±s.d., n = 3 biological replicates. i. Structure models showing how the PRPF3 P9 Lys interacts with HLA-B*27:05 Asp at 116 and HLA-B*27:09 His at 116. The peptide Lys 9 is colored magenta, Asp 116 is colored green, and His 116 is colored cyan. j. AS8.4 T cells incubated with YEIH-pulsed K562 cells, express HLA-B*27:05 or HLA-B*27:09, respectively, were tested for CD69 up-regulation. Data are shown as mean±s.d., n = 3 biological replicates. k. AS8.2 T cells incubate with YEIH-pulsed K562 cells, expressing HLA-B*27:05 or HLA-B*27:09, respectively, were tested for CD69 up-regulation. Data are shown as mean±s.d., n = 3 biological replicates. l. Structure models showing how the YEIH P9 Phe interacts with HLA-B*27:05 Asp at 116 and HLA-B*27:09 His at 116. The peptide Phe 9 is colored magenta, Asp 116 is colored green, and His 116 is colored cyan. m. AU2 T cells incubate with MPP4-pulsed K562 cells, expressing HLA-B*27:05 or HLA-B*27:09, respectively, were tested for CD69 up-regulation. Data are shown as mean±s.d., n = 3 biological replicates. n. Structure models showing how the MPP4 P9 His interacts with HLA-B*27:05 Asp at 116 and HLA-B*27:09 His at 116. o. Thermal stability of refolded peptide/HLA-B*27 complex as determined by differential scanning fluorimetry. The melting temperature, Tm, of each complex is plotted as a bar graph with mean±s.d. (n = 3 technical triplicates).

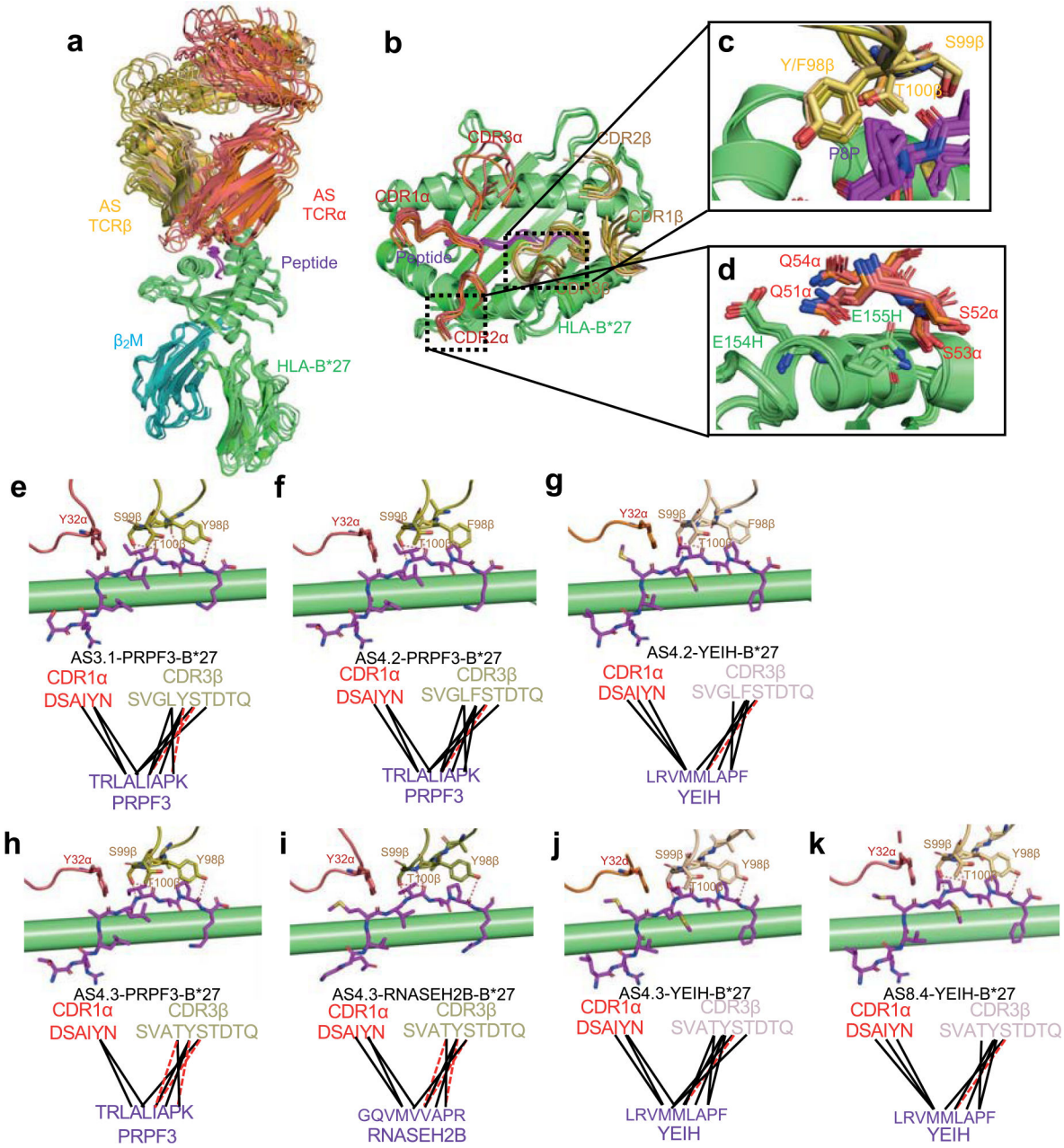


Extended Data Fig. 8 |. SPR sensorgram of AS TCRs and peptide-HLA-B*27:05.

(a-p) Binding analysis of AS TCRs AS3.1, AS4.1, AS4.2, AS4.3 and AS8.4 to GPER1/PRPF3/RNASEH2B/YEIH-HLA-B*27:05 molecules. Increasing concentrations of each TCR was injected over immobilized GPER1/PRPF3/RNASEH2B/YEIH-HLA-B*27:05 complexes. The response unit vs concentration plots were fitted to a steady state affinity model. The SPR measurements were performed once for each TCR-pMHC interaction.



Extended Data Fig. 9 |. Omit maps of AS TCR-peptide-HLA-B*27:05 complex contoured at 1 σ.
a. Simulated annealing composite omit map of the entire AS4.3 TCR-RNASEH2B-HLA-B*27:05 complex contoured at 1 σ. The TCRα chain is colored red; the TCRβ chain is colored yellow; the peptide is colored magenta; the β_2M is colored cyan; the HLA-B*27:05 heavy chain is colored green. (b-h) Simulated annealing composite omit maps of the AS4.3 TCR-RNASEH2B-HLA-B*27:05, AS3.1 TCR-PRPF3-HLA-B*27:05, AS4.2 TCR-PRPF3-HLA-B*27:05, AS4.2 TCR-YEIH-HLA-B*27:05, AS4.3 TCR-PRPF3-HLA-B*27:05, AS4.3 TCR-YEIH-HLA-B*27:05 and AS8.4 TCR-YEIH-HLA-B*27:05 complex interfaces contoured at 1 σ. The TCR CDR1α is colored red; the TCR CDR3β is colored yellow; the peptide is colored magenta; the HLA-B*27:05 heavy chain is colored green.



Extended Data Fig. 10 |. Detailed interactions between AS TCR and peptide-HLA-B*27:05.

a. Side view of superimposed AS TCR-peptide-HLA-B*27:05 complexes. b. Top view of superimposed interface between AS TCR CDR loops and peptide-HLA-B*27:05. c. Superimposed complex structures reveal CDR3β interaction with peptides and HLA-B*27:05. d. Superimposed complex structures reveal CDR2β interaction with HLA-B*27:05. (e-k) Peptide recognition by CDR1α and CDR3β in AS TCR-peptide-HLA-B*27:05 complexes. The upper panels are 'Sticks' indicating TCR residues and peptide residues making contacts. CDR1α is colored red or orange and depicted in cartoon. CDR3β is colored yellow or wheat and depicted in cartoon. Peptide is colored magenta and HLA-B*27:05 is colored green and depicted in cartoon. The lower panels show CDR1α, CDR3β

and peptide sequences and atomic interactions. The black lines indicate van der Waals contacts; red dashed lines indicate hydrogen bonds.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The study received support from the National Institutes of Health (NIH; 5R01AI103867 (K.C.G.) and U19AI057229 (K.C.G.)), the Howard Hughes Medical Institute (K.C.G.), the UK Medical Research Council (MR/M019837/1; G.M.G., A.J.M. and P.B.), Oxford University McMichael Trust Fund (G.M.G and L.I.G), the Rosetrees Trust (M455; G.M.G.), Oxford University John Fell Fund and Medical Sciences Division Internal Funds (G.M.G. and L.I.G.), the UK National Institute for Health Research Oxford Biomedical Research Centre (BRC) (P.B.), the Ministry of Science and Higher Education of the Russian Federation (075-15-2019-1789; E.A.K., I.V.Z. and D.M.C.), the Rheumatology Research Foundation (M.A.P.), the Arthritis National Research Foundation (M.A.P.), the Rheumatic Diseases Research Resource-based Center at Washington University (NIH P30-AR073752; M.A.P. and W.M.Y.), the Bursky Center for Human Immunology and Immunotherapy Programs (W.M.Y.) and the Barnes-Jewish Hospital Foundation (W.M.Y.). The views expressed are those of the author(s) and not necessarily those of the NHS, the National Institute for Health Research or the Department of Health. We thank the Oxford Genomics Centre at the Wellcome Centre for Human Genetics (Wellcome Trust grant reference 203141/Z/16/Z) for the generation and initial processing of sequencing data and J. Webber (Cell Sorting Facility, Kennedy Institute of Rheumatology, University of Oxford) for conducting cell sorting. We also thank N. Ternette, W. Paes and R. Inman for data analysis advice and scientific discussions. The Berkeley Center for Structural Biology is supported in part by the Howard Hughes Medical institute. The Advanced Light Source is a Department of Energy Office of Science User Facility under Contract No. DE-AC02-05CH11231. The ALS-ENABLE beamlines are supported in part by the NIH, National institute of General Medical Sciences (grant P30 GM124169). The schematic cartoon figures in Fig. 1 and Extended Data Fig. 1 were created with BioRender.com.

Our findings support the arthritogenic peptide hypothesis and suggest that microorganisms could elicit autoimmune responses restricted by HLA-B*27.

Data availability

The protein sequences used for target prediction were obtained from Uniprot (http://www.uniprot.org/). The human proteome ID is UP000005640. The *Chlamydia* proteome ID is UP000049987. The *Salmonella* proteome ID is UP000000541. The *Shigella* proteome ID is UP000001006. The *Klebsiella* proteome ID is UP000019183. The *Yersinia* proteome ID is UP000000815. The processed and raw deep-sequencing data have been deposited in the GEO database under the accession code GSE215018. The structures have been deposited in the RCSB protein data bank with the accession codes 7N2N, 7N2O, 7N2P, 7N2Q, 7N2R, 7N2S and 8CX4. There is no restriction on data availability. Source data are provided with this paper.

References

1. Bowness P HLA-B27. Annu. Rev. Immunol. 33, 29–48 (2015). [PubMed: 25861975]
2. Faham M et al. Discovery of T cell receptor beta motifs specific to HLA-B27-positive ankylosing spondylitis by deep repertoire sequence analysis. Arthritis Rheumatol. 69, 774–784 (2017). [PubMed: 28002888]
3. Hanson AL et al. Altered repertoire diversity and disease-associated clonal expansions revealed by T cell receptor immunosequencing in ankylosing spondylitis patients. Arthritis Rheumatol. 72, 1289–1302 (2020). [PubMed: 32162785]

4. Komech EA et al. CD8+ T cells with characteristic T cell receptor beta motif are detected in blood and expanded in synovial fluid of ankylosing spondylitis patients. Rheumatology 57, 1097–1104 (2018). [PubMed: 29481668]
5. Schlosstein L et al. High association of an HL-A antigen, W27, with ankylosing spondylitis. N. Engl. J. Med. 288, 704–706 (1973). [PubMed: 4688372]
6. Brewerton DA et al. Acute anterior uveitis and HL-A 27. Lancet 302, 994–996 (1973). [PubMed: 4127279]
7. Brewerton DA et al. Reiter's disease and HL-A 27. Lancet 302, 996–998 (1973). [PubMed: 4127280]
8. Rosenbaum JT & Asquith M The microbiome and HLA-B27-associated acute anterior uveitis. Nat. Rev. Rheumatol. 14, 704–713 (2018). [PubMed: 30301938]
9. Robinson PC et al. Genetic dissection of acute anterior uveitis reveals similarities and differences in associations observed with ankylosing spondylitis. Arthritis Rheumatol. 67, 149–151 (2015).
10. Finch M et al. Epidemic Reiter's syndrome following an outbreak of shigellosis. Eur. J. Epidemiol. 2, 26–30 (1986). [PubMed: 3533609]
11. Aho K et al. HL-A 27 in reactive arthritis. A study of yersinia arthritis and Reiter's disease. Arthritis Rheumatol. 17, 521–526 (1974).
12. Benjamin R & Parham P Guilt by association: HLA-B27 and ankylosing spondylitis. Immunol. Today 11, 137–142 (1991).
13. Hammer RE et al. Spontaneous inflammatory disease in transgenic rats expressing HLA-B27 and human beta 2m: an animal model of HLA-B27-associated human disorders. Cell 63, 1099–1112 (1990). [PubMed: 2257626]
14. Tran TM et al. Additional human beta2-microglobulin curbs HLA-B27 misfolding and promotes arthritis and spondylitis without colitis in male HLA-B27-transgenic rats. Arthritis Rheumatol. 54, 1317–1327 (2006).
15. Baggia S et al. A novel model of bacterially-induced acute anterior uveitis in rats and the lack of effect from HLA-B27 expression. J. Investig. Med. 45, 295–301 (1997).
16. Allen RL et al. Cutting edge: HLA-B27 can form a novel beta 2-microglobulin-free heavy chain homodimer structure. J. Immunol. 1, 5045–5048 (1999).
17. DeLay ML et al. HLA-B27 misfolding and the unfolded protein response augment interleukin-23 production and are associated with Th17 activation in transgenic rats. Arthritis Rheumatol. 60, 2633–2643 (2009).
18. Kollnberger S et al. Cell-surface expression and immune receptor recognition of HLA-B27 homodimers. Arthritis Rheumatol. 46, 2972–2982 (2002).
19. Bowness P et al. Th17 cells expressing KIR3DL2+ and responsive to HLA-B27 homodimers are increased in ankylosing spondylitis. J. Immunol. 186, 2672–2680 (2011). [PubMed: 21248258]
20. Zheng M et al. TCR repertoire and CDR3 motif analyses depict the role of αβ T cells in ankylosing spondylitis. EBioMedicine 47, 414–426 (2019). [PubMed: 31477563]
21. Duchmann R et al. HLA-B27-restricted cytotoxic T lymphocyte responses to arthritogenic enterobacteria or self-antigens are dominated by closely related TCRBV gene segments. A study in patients with reactive arthritis. Scand. J. Immunol. 43, 101–108 (1996). [PubMed: 8560188]
22. Dulphy N et al. Common intra-articular T cell expansions in patients with reactive arthritis: identical beta-chain junctional sequences and cytotoxicity toward HLA-B27. J. Immunol. 162, 3830–3839 (1999). [PubMed: 10201900]
23. May E et al. TCRB junctional regions from HLA-B27-restricted T cells and HLA-B27 binding peptides display conserved hydropathy profiles in the absence of primary sequence homology. Int. Immunol. 8, 1815–1823 (1996). [PubMed: 8943577]
24. Evans DM et al. Genome-wide association study of ankylosing spondylitis identifies non-MHC susceptibility loci. Nat. Genet. 42, 123–127 (2010). [PubMed: 20062062]
25. York IA et al. Endoplasmic reticulum aminopeptidase 1 (ERAP1) trims MHC class I-presented peptides in vivo and plays an important role in immunodominance. Proc. Natl Acad. Sci. USA 103, 9202–9207 (2006). [PubMed: 16754858]

26. York IA et al. The ER aminopeptidase ERAP1 enhances or limits antigen presentation by trimming epitopes to 8–9 residues. Nat. Immunol. 3, 1177–1184 (2002). [PubMed: 12436110]
27. Martin-Esteban A, Gomez-Molina P, Sanz-Bravo A & Lopez de Castro JA Combined effects of ankylosing spondylitis-associated ERAP1 polymorphisms outside the catalytic and peptide-binding sites on the processing of natural HLA-B27 ligands. J. Biol. Chem. 289, 3978–3990 (2014). [PubMed: 24352655]
28. Diaz-Pena R, Lopez-Vazquez A & Larrea L Old and new HLA associations with ankylosing spondylitis. Tissue Antigens 80, 205–213 (2012). [PubMed: 22881057]
29. Colmegna I, Cuchacovich R & Espinoza LR HLA-B27-associated reactive arthritis: pathogenetic and clinical considerations. Clin. Microbiol. Rev. 17, 348–369 (2004). [PubMed: 15084505]
30. Ebringer A, Rashid T, Tiwana H & Wilson C A possible link between Crohn's disease and ankyloisng spondylitis via Klebsiella infections. Clin. Rheumatol. 26, 289–297 (2007). [PubMed: 16941202]
31. Rashid T, Wilson C & Ebringer A The link between ankylosing spondylitis, Crohn's disease, Klebsiella, and starch consumption. Clin. Dev. Immunol. 2013, 872632 (2013). [PubMed: 23781254]
32. Stewart-Jones GBE et al. Crystal structures and KIR3DL1 recognition of three immunodominant viral peptides complexed to HLA-B*2705. Eur. J. Immunol. 35, 341–351 (2005). [PubMed: 15657948]
33. Hulsmeyer M et al. HLA-B27 subtypes differentially associated with disease exhibited subtle structural alterations. J. Biol. Chem. 277, 47844–47853 (2002). [PubMed: 12244049]
34. Lim Kam Sian TCC et al. Allelic association with ankylosing spondylitis fails to correlate with human leukocyte antigen B27 homodimer formation. J. Biol. Chem. 294, 20185–20195 (2019). [PubMed: 31740583]
35. Cole DK et al. Human TCR-binding affinity is governed by MHC class restriction. J. Immunol. 178, 5727–5734 (2007). [PubMed: 17442956]
36. Petersen J et al. Diverse T cell receptor gene usage in HLA-DQ8-associated celiac disease converges into a consensus binding solution. Structure 24, 1643–1657 (2016). [PubMed: 27568928]
37. Yang X et al. Structural basis for clonal diversity of the human T-cell response to a dominant influenza virus epitope. J. Biol. Chem. 292, 18618–18627 (2017). [PubMed: 28931605]
38. May E et al. Conserved TCR beta chain usage in reactive arthrits; evidence for selection by a putative HLA-B27-associated autoantigen. Tissue Antigens 60, 299–308 (2002). [PubMed: 12472659]
39. Broughton SE et al. Biased T cell receptor usage directed against human leukocyte antigen DQ8-restricted gliadin peptides is associated with celiac disease. Immunity 37, 611–621 (2012). [PubMed: 23063329]
40. Venturi V et al. Sharing of T cell receptors in antigen-specific responses is driven by convergent recombination. Proc. Natl Acad. Sci. USA 103, 18691–18696 (2006). [PubMed: 17130450]
41. Roldan EQ et al. Different TCRBV genes generate biased pattern of V-D-J diversity in human T cells. Immunogenetics 41, 91–100 (1995). [PubMed: 7806301]
42. Schittenhelm RB et al. Revisiting the arthritogenic peptide theory: quantitative not qualitative changes in the peptide repertoire of HLA-B27 allotypes. Arthritis Rheumatol. 67, 702–713 (2015). [PubMed: 25418920]
43. Barnea E et al. The human leukocyte antigen (HLA)-B27 peptidome in vivo, in spondyloarthritis-susceptible HLA-B27 transgenic rats and the effect of Erap1 deletion. Mol. Cell Proteomics 16, 642–662 (2017). [PubMed: 28188227]
44. Garcia-Medal N et al. Peptide handling by HLA-B27 subtypes influences their biological behavior, association with ankylosing spondylitis and susceptibility to endoplasmic reticulum aminopeptidase 1 (ERAP1). Mol. Cell. Proteomics 13, 3367–3380 (2014). [PubMed: 25187574]
45. Schittenhelm RB et al. Human leukocyte antigen (HLA)-B27 allotype-specific binding and candidate arthritogenic peptides revealed through heuristic clustering of data-independent acquisition mass spectrometry (DIA-MS) data. Mol. Cell. Proteomics 15, 1867–1876 (2016). [PubMed: 26929215]

46. Chagin AS & Sävendahl L GPR30 estrogen receptor expression in the growth plate declines as puberty progresses. J. Clin. Endocrinol. Metab. 92, 4873–4877 (2007). [PubMed: 17878253]
47. Heino TJ, Chagin AS & Sävendahl L The novel estrogen receptor G-protein-coupled receptor 30 is expressed in human bone. J. Endocrinol. 197, R1–R6 (2008). [PubMed: 18434348]
48. Gautam P et al. Multi-species single-cell transcriptomic analysis of ocular compartment regulons. Nat. Commun. 28, 5675 (2021).
49. Cole DK et al. Hotspot autoimmune T cell receptor binding underlies pathogen and insulin peptide cross-reactivity. J. Clin. Invest. 26, 2191–2204 (2016).
50. Petersen J et al. T cell receptor cross-reactivity between gliadin and bacterial peptides in celiac disease. Nat. Struct. Mol. Biol. 27, 49–61 (2020). [PubMed: 31873306]
51. Standardization of Uveitis Nomenclature (SUN) Working Group. Classification criteria for spondyloarthritis/HLA-B27-associated anterior uvetis. Am. J. Ophthalmol. 228, 117–125 (2021). [PubMed: 33845004]
52. Rudwaleit M et al. The development of Assessment of SpondyloArthritis International Society classification criteria for axial spondyloarthritis (part II): validation and final selection. Ann. Rheum. Dis. 68, 777–783 (2009). [PubMed: 19297344]
53. Jabs DA et al. Standardization of uveitis nonmenclature for reporting clinical data. Results of the first international workshop. Am. J. Ophthalmol. 140, 509–516 (2005). [PubMed: 16196117]
54. Hassman LM et al. Clinicomolecular identification of conserved and individualized features of granulomatous uveitis. Ophthalmol. Sci. 10.1016/j.xops.2021.100010 (2021).
55. Pogorelyy MV et al. Persisting fetal clonotypes influence the structure and overlap of adult human T cell receptor repertoires. PLoS Comput. Biol. 13, e1005572 (2017). [PubMed: 28683116]
56. Shugay M et al. Towards error-free profiling of immune repertoires. Nat. Methods 11, 653–655 (2014). [PubMed: 24793455]
57. Bolotin DA et al. MiXCR: software for comprehensive adaptive immunity profiling. Nat. Methods 12, 380–381 (2015). [PubMed: 25924071]
58. Bowness P, Allen RL & McMichael AJ Identification of T cell receptor recognition residues for a viral peptide presented by HLA B27. Eur. J. Immunol. 24, 2357–2363 (1994). [PubMed: 7523137]
59. Lybarger L et al. Enhanced immune presentation of a single-chain major histocompatibility complex class I molecule engineered to optimize linkage of a C-terminally extended peptide. J. Biol. Chem. 278, 27105–27111 (2003). [PubMed: 12732632]
60. Birnbaum ME et al. Deconstructing the peptide-MHC specificity of T cell recognition. Cell 157, 1073–1087 (2014). [PubMed: 24855945]
61. Garboczi DN et al. Structure of the complex between human T-cell receptor, viral peptide and HLA-A2. Nature 384, 134–141 (1996). [PubMed: 8906788]
62. Aricescu AR, Lu W & Jones EY A time- and cost-efficient system for high-level protein production in mammalian cells. Acta Crystallogr. D 62, 1243–1250 (2006). [PubMed: 17001101]
63. Ellis SA, Taylor C & McMichael A Recognition of HLA-B27 and related antigen by a monoclonal antibody. Hum. Immunol. 5, 49–59 (1982). [PubMed: 6981636]
64. Garboczi DN, Hung DT & Wiley DC HLA-A2-peptide complexes: refolding and crystallization of molecules expressed in *Escherichia coli* and complexed with single antigenic peptides. Proc. Natl Acad. Sci. USA 89, 3429–3433 (1992). [PubMed: 1565634]
65. Emsley P et al. Features and development of COOT. Acta Crystallogr. D 66, 486–501 (2010). [PubMed: 20383002]
66. Liebschner D et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. Acta Crystallogr. D 75, 861–877 (2019).

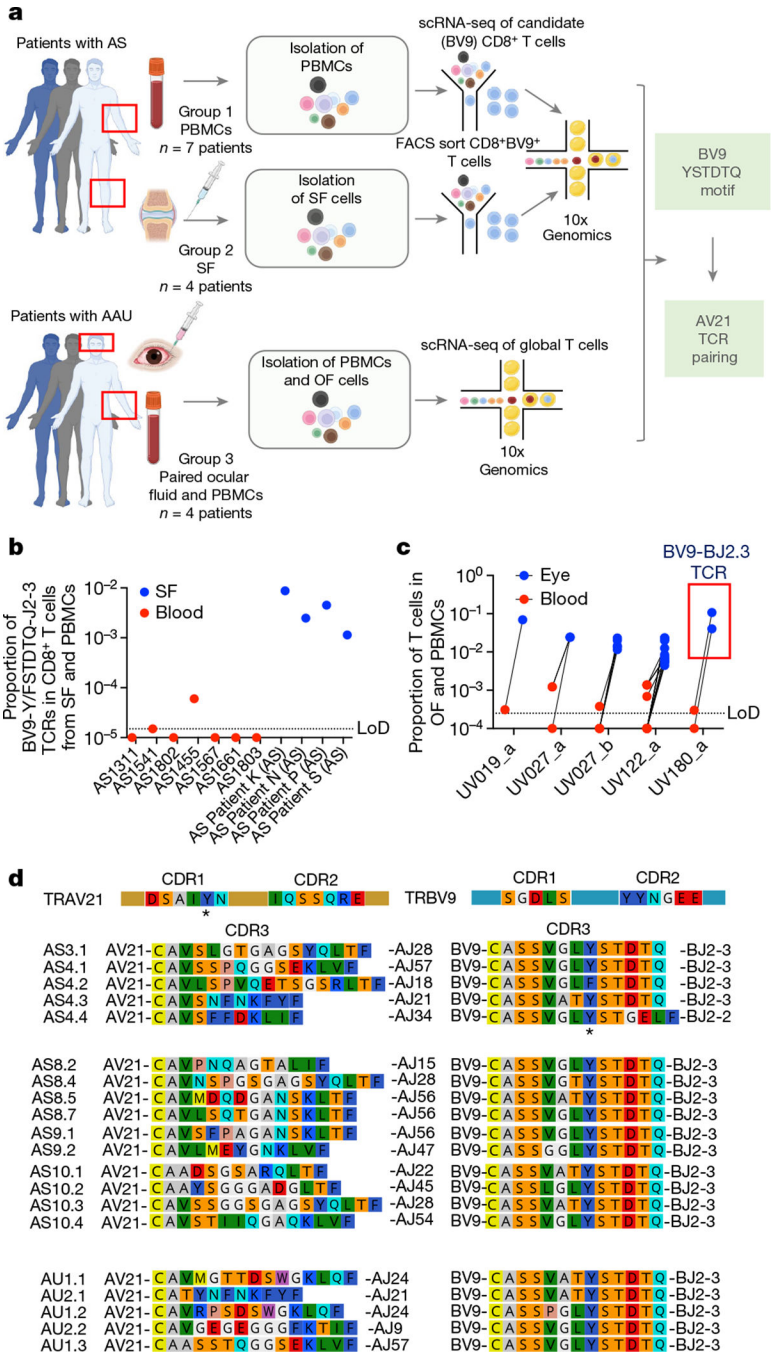


Fig. 1 |. Identification of AS- and AAU-associated TRBV9–TRBJ2.3 TCRs.

a, Schematic of TCR isolation from PBMCs and synovial fluid (SF) cells from patients with AS, and from PBMCs and ocular fluid cells from patients with AAU. Created with BioRender.com. FACS, fluorescence-activated cell sorting; OF, ocular fluid. **b**, The frequency of $CD8^+$ T cells containing the YSTDTQ TRBJ2.3 motif from sorted $BV9^+CD8^+$ T cells in PBMCs and synovial fluid from patients with AS. The level of detection (LoD) of blood samples is indicated by a dashed line and is the median proportion of a singleton clonotype from all blood samples. **c**, Unbiased single-cell TCR sequencing of ocular fluid

cells and PBMCs from patients with AAU. The proportions of expanded ocular TCR clonotypes (defined as greater than 10 barcodes in the eye) are compared between the ocular fluid and corresponding PBMCs from patients with AAU. The '_a' and '_b'- suffixes designate the first and second samplings, respectively. The level of detection of blood samples is indicated by a dashed line and is the median proportion of a singleton clonotype from all blood samples. **d**, Partial amino acid sequence alignment of TCRs derived from PBMCs and synovial fluid from patients with AS and TCRs derived from PBMCs and ocular fluid from patients with AAU used in this study, in RasMol colouring. The invariant CDR1 and CDR2 sequences are shown for TRAV21 and TRBV9 in the top row. CDR3 amino acid sequences are shown in the bottom rows along with corresponding variable and joining gene usage. Structurally important bulky residues are marked with asterisks.

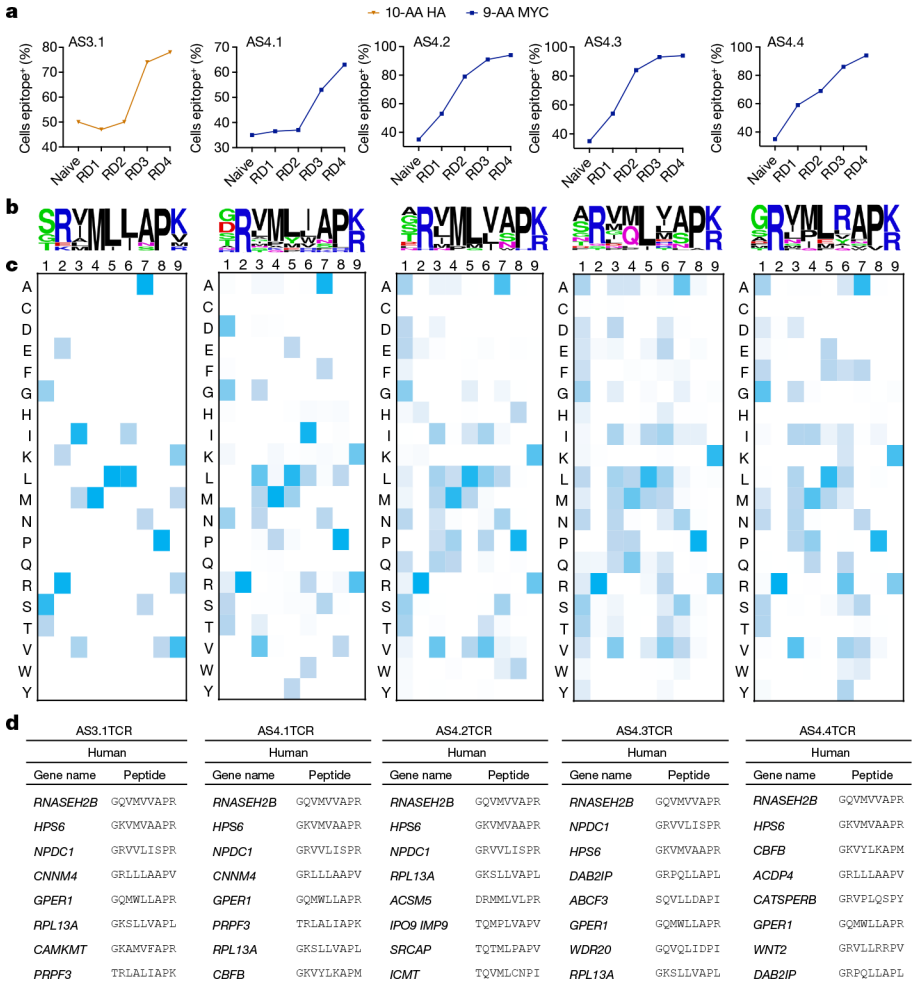


Fig. 2 |. Screening TRBV9-TRBJ2.3 TCRs on HLA-B*27:05 yeast display libraries.

a, AS3.1, AS4.1, AS4.2, AS4.3 and AS4.4 TCRs selected on HLA-B*27:05 libraries. Enrichments from the naive library to the fourth round (RD4) were measured through flow cytometry and are plotted as the percentage of cells that are epitope positive. **b**, WebLogos representing the unique fourth-round-selected peptides for each AS TCR based on deep-sequencing reads. The size of each amino acid letter represents its abundance at the given position among the unique peptides. **c**, Heatmap plots showing the amino acid composition per position of the peptides enriched after the fourth round of selection. A darker colour represents a greater abundance of a given amino acid at a specific position. **d**, Top algorithm-predicted peptides from the human proteome based on AS3.1, AS4.1, AS4.2, AS4.3 and AS4.4 deep-sequencing data.

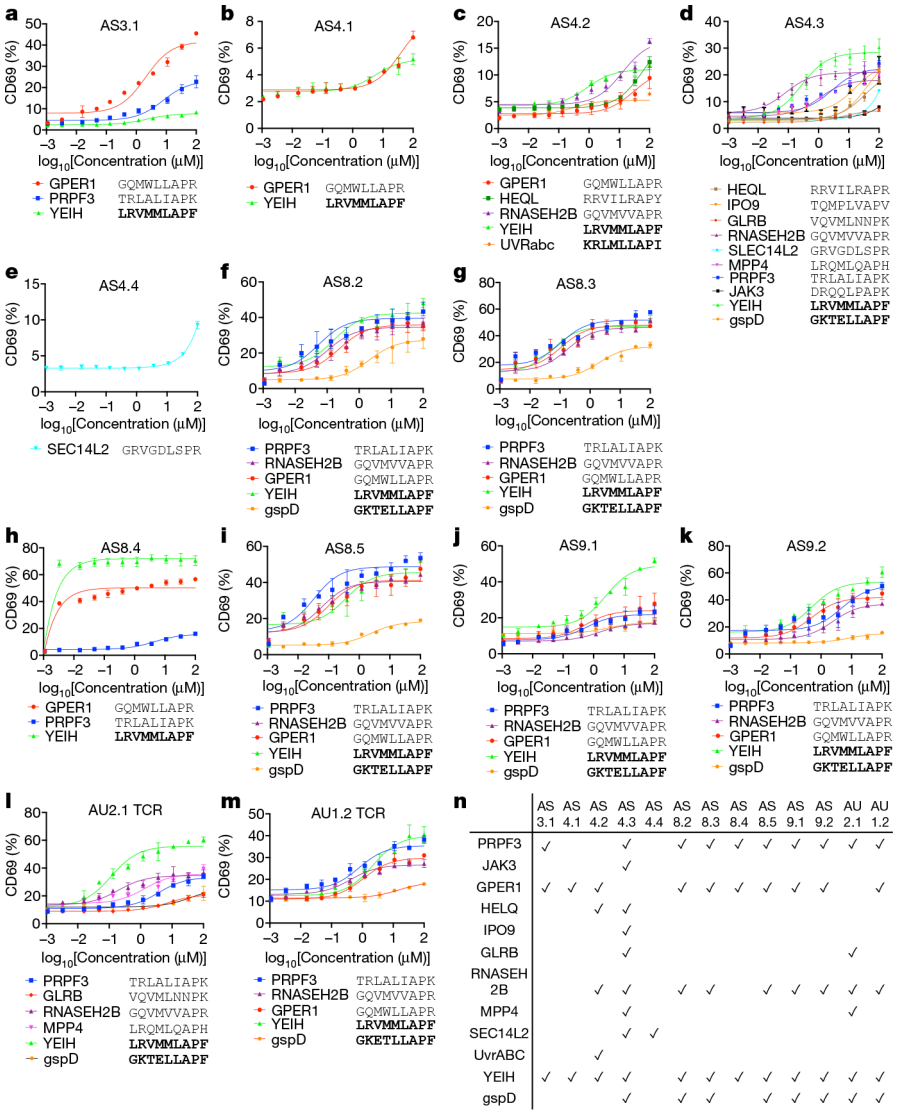
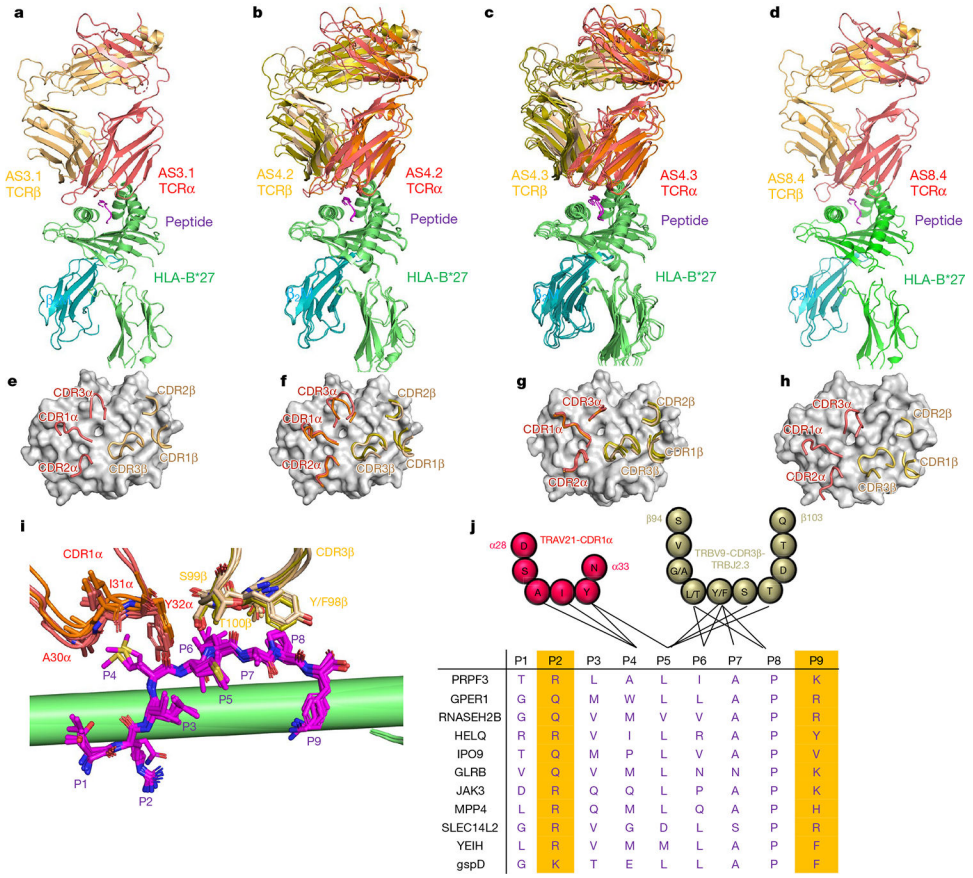


Fig. 3 |. Validation of TCR binding and activation by predicted human and microbial antigens. **a**–**m**, TCRs from patients with AS or AAU were transduced into CD8$^+$ SKW-3 cells and sorted for stable TCR (IP26) and CD3 (UCHT1) co-expression. Wild-type HLA-B*27:05 was transduced into K562 cells, and the cells were sorted for stable HLA molecule expression. APC lines were pulsed with 1:3 serial-diluted peptides starting from 100 μM for 2 h, co-incubated with the T cell lines for 18 h and analysed for CD69 expression by flow cytometry. AS3.1 (**a**), AS4.1 (**b**), AS4.2 (**c**), AS4.3 (**d**), AS4.4 (**e**), AS8.2 (**f**), AS8.3 (**g**), AS8.4 (**h**), AS8.5 (**i**), AS9.1 (**j**), AS9.2 (**k**), AU2.1 (**l**) and AU1.2 (**m**) were tested for CD69 activation. Dose–response curves for each human and common microbial stimulatory peptide are shown with means of biological triplicates with standard error of the mean. The microbial-derived peptides are in bold to differentiate them from human-derived peptides. Representative data are shown as mean ± s.d. (n = 3 biological triplicates). **n**, The matrix of activated peptides corresponding to each TCR included in the peptide activation assay. Check marks indicate activation of TCR by the corresponding peptide, as defined

by CD69 upregulation when dosed with 100 μM (E_{max}) of individual peptide compared to dimethylsulfoxide (DMSO) control (n = 3 biological triplicates).



	P1	P2	P3	P4	P5	P6	P7	P8	P9
PRPF3	T	R	L	A	L	I	A	P	K
GPER1	G	Q	M	W	L	L	A	P	R
RNASEH2B	G	Q	V	M	V	V	A	P	R
HELQ	R	R	V	I	L	R	A	P	Y
IPO9	T	Q	M	P	L	V	A	P	V
GLRB	V	Q	V	M	L	N	N	P	K
JAK3	D	R	Q	Q	L	P	A	P	K
MPP4	L	R	Q	M	L	Q	A	P	H
SLEC14L2	G	R	V	G	D	L	S	P	R
YEIH	L	R	V	M	M	L	A	P	F
gspD	G	K	T	E	L	L	A	P	F

Fig. 4 |. Biophysical and structural basis of AS TCR–peptide–HLA-B*27:05 complexes.

a–**d**, Overview of AS3.1–PRPF3–B*27 (**a**), AS4.2–PRPF3/YEIH–B*27 (**b**), AS4.3–PRPF3/YEIH/RNASEH2B–B*27 (**c**) and AS8.4–YEIH–B*27 (**d**) complexes. The HLA-B*27:05 heavy chain, β_2m, peptide, TCRα chain and TCRβ chain are depicted in cartoon mode. HLA-B*27:05 is coloured green; β_2m is coloured cyan; peptide is coloured magenta; TCRα chain is coloured red; and TCRβ chain is coloured yellow. **e**–**h**, Top view of CDRs from AS3.1–PRPF3–HLA-B*27:05 (**e**), AS4.2–PRPF3/YEIH–HLA-B*27:05 (**f**), AS4.3–PRPF3/YEIH/RNASEH2B–HLA-B*27:05 (**g**) and AS8.4–YEIH–HLA-B*27:05 (**h**) complex crystal structures. Peptide–HLA-B*27:05 is coloured light grey and depicted in surface mode. CDR3α is coloured red and depicted in cartoon mode. CDR3β is coloured yellow and depicted in cartoon mode. **i**, Detailed peptide recognition through AS TCR CDR1α and CDR3β. The TCRα chain is coloured red, the TCRβ chain is coloured yellow, the peptide is coloured magenta, and HLA-B*27:05 is coloured green. **j**, Detailed amino acid interactions between AS TCR and peptides. The interactions between CDR1α, CDR3β and peptides are depicted as black lines. The interaction cutoff is 4 Å.