PNAS
www.pnas.org

# Supplementary Information for

## The $\alpha\beta$TCR mechanosensor exploits dynamic ectodomain allostery to optimize its ligand recognition site

**Wonmuk Hwang, Robert J. Mallis, Matthew J. Lang and Ellis L. Reinherz**

**W.H: hwm@tamu.edu, E.L.R: ellis_reinherz@dfci.harvard.edu**

**This PDF file includes:**

Supplementary text
Figs. S1 to S8
Legends for Movies S1 to S9
SI References

**Other supplementary materials for this manuscript include the following:**

Movies S1 to S9

## Supporting Information Text

## Structure Preparation

All simulation systems were built based on PDB 1OGA. For simulations where load is applied to the complex, terminal strands were added as explained for T$\alpha\beta$-pMHC below. Other systems were constructed with various truncation schemes. N- and C-termini of all polypeptide chains were left charged. Water molecules in the original PDB structure were kept in the simulation system.

**T$\alpha\beta$-pMHC:** In the original PDB 1OGA, MHC ends at P276. We added $^{277}$SSQPTIPI$^{284}$ based on the sequence of UniProt P01892. For TCR$\alpha\beta$, the $\alpha$- and $\beta$-chains are visible up to P202 and Q245, respectively (P202 of $\alpha$-chain and Q245 of $\beta$-chain have only the backbone N atom visible). The missing C-terminal residues had altered sequences for disulfide formation in the structural study (1). We thus added more native sequences, $^{203}$SSCDVKLVEKSFETDT$^{219}$ for $\alpha$-chain, based on sequence in GenBank FJ795359.1, and for $\beta$-chain, we added $^{245}$CGFTSVSYQQGVLSA$^{259}$ based on UniProt A0A087X076 (Q245 was replaced with C245). The added strands serve as spacers when the terminal ends are restrained (Fig. 1A), hence their precise sequence is unlikely to affect the result.

Disulfide bonds were formed between cysteine residues at locations: MHC: 101-164, 203-259, $\beta 2m$: 25-80, TCR$\alpha$: 24-90, 134-184, TCR$\beta$: 25-93, 145-210, and TCR$\alpha$-TCR$\beta$: $\alpha$206-$\beta$245. Protonation sites for histidines were determined to promote hydrogen bond (H-bond) formation with nearby residues, or in case nearby residues are not amenable for H-bond formation, to make the water-facing nitrogen of the imidazole ring to be charged. The N$^{\epsilon}$ atom was protonated, for MHC: 70, 74, 191, 192, 197, 263; $\beta 2m$: 31, 84; and TCR$\beta$: 137, 167, 207. Histidines in other locations were protonated at N$^{\delta}$.

**T$\alpha\beta$:** Since no positional restraints were applied, the last residues were made $\alpha$D207 and $\beta$G246.

**dFG:** In the $\beta$-chain of T$\alpha\beta$, L217–P230 were removed, and G216 and V231 were joined covalently.

**V$\alpha\beta$:** The last residues were $\alpha$N112 and $\beta$E115.

**V$\alpha\beta$-pMHC:** For the V-module, V$\alpha\beta$ was used. For MHC, residues were kept to S277.

**T$\alpha\beta$-pMHC$^{0}$:** T$\alpha\beta$, and MHC of V$\alpha\beta$-pMHC were used. The system did not have added strands since no load was applied.

## MD Simulation

**Initial relaxation of added strands:** After building a T$\alpha\beta$-pMHC system, the added strands take straight conformations. They were conformationally relaxed and the C-terminal ends of TCR $\alpha$- and $\beta$-chains were brought close together by fixing the positions of all atoms except for the added strands, and applying a series of energy minimization and short MD simulation within the FACTS implicit solvent environment (2). Further relaxation of the added strands was performed in the solvated system, as explained below.

**MD simulation protocol:** The CHARMM param36 all-atom force field was used. For simulations with positional restraints (V$\alpha\beta$-pMHC, T$\alpha\beta$-pMHC$^{\mathrm{low,high,0}}$ and dFG-pMHC$^{h1,h2}$), an elongated orthorhombic solvation box was used. The size of the box was such that atoms of the protein (including hydrogens) are at least 13 Å away from the nearest face of the box. For simulations without positional restraints, a cubic box was used with the side length 24 Å larger than the maximum span of the protein. The water box was filled with pre-equilibrated TIP3P water boxes, and water molecules whose oxygen atom is less than 2.8 Å from the nearest protein heavy atom were removed. The system was made electrically neutral by replacing randomly chosen water molecules with sodium and chloride ions to a concentration of approximately 50 mM. We performed three independent neutralization attempts that differ in locations of added ions. For each, a brief energy minimization was performed and the configuration with the lowest energy was used. Representative sizes of simulation systems are, for T$\alpha\beta$-pMHC$^{\mathrm{high}}$, 231.2×91.9×98.7 Å$^3$ containing 212,498 atoms, and for T$\alpha\beta$ (cubic system), 94.7 Å$^3$ containing 75,584 atoms.

After neutralization, a 5-stage energy minimization procedure was applied. For the first stage, protein backbone and side chain heavy atoms were harmonically restrained with spring constants of 40 kcal/[mol·Å$^2$] and 20 kcal/[mol·Å$^2$], respectively. Energy minimization was applied with 200 steps of steepest descent method followed by 200 steps of adopted-basis Newton-Raphson method. In the second stage, spring constants were reduced by 10 kcal/[mol·Å$^2$] (backbone heavy atoms) and 5 kcal/[mol·Å$^2$] (side chain heavy atoms), and the same energy minimization protocol was applied. This was repeated until the 4th stage using progressively weaker constraints. At the 5th stage of energy minimization, no restraint was applied.

The energy minimized system was heated and equilibrated with a 5-kcal/[mol·Å$^2$] harmonic constraint applied to the backbone heavy atoms. Heating was from 30 K to 300 K for 100 ps, and equilibration was at 300 K for 200 ps. A constant temperature and pressure (CPT) thermostat was used under 1 atm. An additional CPT simulation at 1 atm, 300 K was performed for 2 ns with a weak 0.001-kcal/[mol·Å$^2$] harmonic constraint applied to backbone alpha-carbon atoms only. For the T$\alpha\beta$-pMHC complex, no restraints were applied to the terminal added strands so that they can conformationally relax. Subsequent simulations for individual systems are explained below.

For simulation, the particle-mesh Ewald summation method was used to account for the long-range effect of electrostatic interactions. The SHAKE method was applied to fix the length of covalent bonds involving hydrogen atoms, which allowed using a 2-fs integration time step. The cutoff distance for nonbonded interactions was 12 Å. The domain decomposition (DOMDEC) module of CHARMM (3) was used for efficient parallelization of simulation. Coordinates were saved every 20 ps. For a 500-ns simulation as an example, this yields 25,000 coordinate frames, and skipping the initial 200 ns leaves 15,000 frames to analyze.

**Simulation length:** Production runs were carried out under a constant temperature (300 K) and volume (NVT) condition, and they lasted for: 520 ns (T$\alpha\beta$-pMHC$^{\text{low}}$), 560 ns (T$\alpha\beta$-pMHC$^{\text{high}}$), 510 ns (T$\alpha\beta$, V$\alpha\beta$, and V$\alpha\beta$-pMHC), 500 ns (dFG and T$\alpha\beta$-pMHC$^{0}$), 430 ns (dFG-pMHC$^{h1}$), and 380 ns (dFG-pMHC$^{h2}$). The aggregate simulation time was over 4.4 $\mu$s.

**Laddered extension:** For T$\alpha\beta$-pMHC$^{\text{low}}$ and T$\alpha\beta$-pMHC$^{\text{high}}$, we performed a simulation where harmonic restraints were applied with increasing separation. After the 2-ns equilibration run where the added strands underwent conformational relaxation, a 1-kcal/[mol·Å$^2$] harmonic constraint was applied to the alpha-carbon atom of the C-terminal I284 of MHC and the center of mass of the two alpha-carbon atoms of the C-terminal residues of T$\alpha\beta$. For the latter, a flat-bottom distance restraint was also applied to keep the two alpha-carbon atoms within 10 Å. A 2-ns CPT simulation was performed, after which the distance between the harmonic restraints was increased by 8 Å followed by another 2-ns run. In this way, we increased the distance up to 64 Å in 8-Å increments. At each extension, we carried out an additional 60-ns run and examined conformational behaviors and the average force generated. We selected the very first system without any extra extension for T$\alpha\beta$-pMHC$^{\text{low}}$ and the one with the 16-Å extension for T$\alpha\beta$-pMHC$^{\text{high}}$. For the stepwise extension, the longitudinal length of the water box was set to 300 Å, to enclose the maximally extended system. After selecting systems for T$\alpha\beta$-pMHC$^{\text{low}}$ and T$\alpha\beta$-pMHC$^{\text{high}}$, water boxes were truncated to reduce their length to be 26 Å (13 Å on each end) larger than the maximum span of the complex in respective cases.

The initial states of dFG-pMHC$^{h1}$ and dFG-pMHC$^{h2}$ were from the beginning of the simulations for T$\alpha\beta$-pMHC$^{\text{high}}$ and T$\alpha\beta$-pMHC$^{\text{low}}$, respectively. After removing the FG-loop, the system was re-neutralized to compensate for the change in the net charge after the FG-loop deletion. It subsequently went through energy minimization, heating, and equilibration as described above. For dFG-pMHC$^{h1}$, a production run followed. For dFG-pMHC$^{h2}$, laddered extension simulations as explained above were performed in 8-Å increments, each simulation lasting 4 ns. The simulation where the average extension is similar to that for T$\alpha\beta$-pMHC$^{\text{high}}$ was selected for a longer simulation. The separation between the centers of the harmonic restraints was 193.67 Å for dFG-pMHC$^{h1}$ (the same as T$\alpha\beta$-pMHC$^{\text{high}}$), and 192.27 Å for dFG-pMHC$^{h2}$. Note that the extension is slightly smaller for the latter even though it dissociated earlier (Fig. 5 vs. Fig. S6).

**Restraints in other systems:** For V$\alpha\beta$-pMHC, a 0.01-kcal/[mol·Å$^2$] harmonic constraint was applied to the alpha-carbon atoms of the $\alpha$3 domain of MHC (P185–P276), to keep the complex from rotating to become transverse within the elongated simulation box. In T$\alpha\beta$-pMHC$^{0}$, we applied a 0.2-kcal/[mol·Å$^2$] harmonic constraint to each alpha-carbon atom within $\alpha$3 of MHC that had less than 0.5-Å root-mean-square fluctuation (RMSF) in T$\alpha\beta$-pMHC$^{\text{low}}$, which were residues P185–H188, C203–F208, I213–L215, Q242–V247, and T258–H263. More refined restraints were possible since we performed simulation of T$\alpha\beta$-pMHC$^{0}$ after analyzing other systems.

## Analysis

**RMSF calculation:** We calculated RMSF of alpha-carbon atoms separately for each domain, after the initial 200 ns. The RMSF is thus unaffected by the relative motion among different subdomains.

**Contact analysis:** Identification of H-bonds and nonpolar contacts was adopted from the method we developed earlier (4). Briefly, H-bonds were identified with the donor-acceptor distance cutoff of 2.4 Å. Nonpolar contacts were identified with a 3.0-Å cutoff distance between neutral atoms (absolute value of partial charge less than 0.3$e$; $e = 1.6 \times 10^{-19}$ C). The average occupancy of a contact is the fraction of frames over which the bond is formed during the measurement period (typically after the initial 200 ns), and the instantaneous occupancy is the rolling (running) average with a 0.8-ns (40 frames) window. Occupancy heat maps (e.g., Fig. 2A,B) were generated using Python.

To calculate the Hamming distance, we identified the initial high-occupancy contacts as those with greater than 80% average occupancy during the first 50 ns. At a given frame, the Hamming distance was defined as the number of these contacts that are absent.

**PCA for V$\alpha$-V$\beta$ motion:** To assign a triad for V$\alpha$, alpha-carbon atoms in the central four $\beta$-strands that have less than 0.5-Å RMSF in T$\alpha\beta$-pMHC$^{\text{low}}$ were used, which are L20–N25 ($b$-strand; strand labels follow the immunoglobulin fold nomenclature (5)), Q34–E39 ($c$-strand), S73–T78 ($e$-strand), and G86–A91 ($f$-strand). They belong to the structural core of the immunoglobulin fold (5). For equal weighing, 6 atoms from each strand were used. Their center of mass was used as the centroid of the triad for V$\alpha$. The major axis of the least-square fitting plane of these atoms was assigned as the arm $\mathbf{e_3}$, which points along the direction of the $\beta$-strands (Fig. 3A). In the structural core of the immunoglobulin fold, the $b$-strand flanks the $e$-strand, and together they belong to the outer $\beta$-sheet, and they face the $c$- and $f$-strands on the inner $\beta$-sheet. The $\mathbf{e_1}$ arm of the triad was assigned as a unit vector perpendicular to $\mathbf{e_3}$ and lies on the plane spanned by $\mathbf{e_3}$ and a vector pointing from the center of mass of the selected atoms in the inner $\beta$-sheet to that in the outer $\beta$-sheet (Fig. 3A). This fixes $\mathbf{e_2}$ as $\mathbf{e_2} = \mathbf{e_3} \times \mathbf{e_1}$. Triad for V$\beta$ was assigned similarly, using alpha-carbon atoms of V21–Q26 ($b$-strand), Y35–D40 ($c$-strand), F76–T81 ($e$-strand), and F90–S95 ($f$-strand).

Coordinate frames were aligned relative to the alpha-carbon atoms explained above, and triads were assigned at each frame. PCA was performed for frames after the initial 200 ns. After subtracting the average triads from the triads of each frame, singular-value decomposition was performed using a custom-written FORTRAN95 program that calls the Linear Algebra PACKage (LAPACK) (6). The three arms of each triad are orthogonal unit vectors, so the amplitude of a principal component (PC) refers to the rotation of these arms in units of radians (Fig. 3D). However, since the amplitude is for the motion of the 6 arms, it does not translate into a linear rotation angle.

**PCA for V-C motion:** Triads (centroids and 3 arms) for V$\alpha$ and V$\beta$ were assigned as explained above. Residues for hinges (Fig. 4C) were alpha-carbon atoms with large RMSF (in T$\alpha\beta$-pMHC$^{\mathrm{low}}$), which were Q114 for the $\alpha$-chain, and D116 and K118 for the $\beta$-chain. The latter two had almost identical RMSF, 1.40 Å, and their center of mass was used as the hinge of the $\beta$-chain. Alpha-carbon atoms used for the C-module were, for C$\alpha$, A119–R124 (*a*-strand), V133–D138 (*b*-strand), V153–K158 (*d*-strand), and S172–D177 (*e*-strand); and for C$\beta$, T142–A147 (*b*-strand), E156–V161 (*c*-strand), S191–V196 (*e*-strand), and F208–Q213 (*f*-strand). To build a bead-on-chain (BOC) model, a "bead" was assigned to the center of mass of these atoms in each domain (the 4 sub-domains, and two hinges; Fig. 4C). For the V-module, locations of beads coincide with centroids of triads.

Coordinate frames were aligned relative to the alpha-carbon atoms used for assigning beads for the C-module. V-module triads and BOC were constructed using each of the aligned coordinate frames. In a BOC model, the position of a bead represents the location of the corresponding domain in units of Ångströms, whereas triads are orientational markers of the V-module. We thus applied PCA separately for the BOC model and the triads. With the C-module as the reference of alignment, PCs for the triads is a mixture of both the V$\alpha$-V$\beta$ relative motion as well as the rotation of the V-module as a whole, where the latter has a greater contribution (translation of the triads relative to the C-module is eliminated by subtracting the average triads). We thus did not use PCs for the triads other than PC1 for illustrating the V-C bending motion (blue arrows on the arms of triads in Fig. 4B). PCs for the BOC inform the motion of the V-module relative to the C-module, as explained in Results.

**CDR3 distance:** To measure the CDR3 distance, 5 alpha-carbon atoms of each CDR3 loop were used: G94–N98 (CDR3$\alpha$) and S97–Y101 (CDR3$\beta$). The CDR3 distance is the center of mass distance between these two groups of atoms.

**Movie S1. Overview of T$\alpha\beta$-pMHC$^{\mathrm{low}}$ and T$\alpha\beta$-pMHC$^{\mathrm{high}}$. 0–520 ns interval is shown. Transparent pink dots: Center of mass trajectory of the antigenic peptide. Instantaneous forces applied at the ends by the harmonic restraints are shown as blue arrows, with length proportional to the magnitude.**

**Movie S2. Schematic illustration of 3 PCA modes of the V$\alpha$-V$\beta$ motion. Spheres in ice blue color denote CDR3 loops.**

**Movie S3. Illustration of PC1 of the V$\alpha$-V$\beta$ triads. Structure at the beginning of each simulation is overlaid.**

**Movie S4. Illustration of PC2 of the V$\alpha$-V$\beta$ triads. Structure at the beginning of each simulation is overlaid.**

**Movie S5. Illustration of PC3 of the V$\alpha$-V$\beta$ triads. Structure at the beginning of each simulation is overlaid.**

**Movie S6. Schematic illustration of the 3 PCA modes of the V-C motion.**

**Movie S7. Illustration of PC1–PC3 of the V-C BOC. Structure at the beginning of each simulation is overlaid.**

**Movie S8. dFG-pMHC$^{h1}$ simulation.**

**Movie S9. dFG-pMHC$^{h2}$ simulation.**

## References


1. GBE Stewart-Jones, AJ McMichael, JI Bell, DI Stuart, EY Jones, A structural basis for immunodominant human T cell receptor recognition. *Nat. immunology* **4**, 657–663 (2003).
2. U Haberthür, A Caflisch, FACTS: Fast analytical continuum treatment of solvation. *J. Comput. Chem.* **29**, 701–715 (2008).
3. AP Hynninen, MF Crowley, New faster CHARMM molecular dynamics engine. *J. Comput. Chem.* **35**, 406–413 (2014).
4. W Hwang, MJ Lang, M Karplus, Kinesin motility is driven by subdomain dynamics. *eLife* **6**, e28948 (2017).
5. P Bork, L Holm, C Sander, The immunoglobulin fold. *J. Mol. Biol.* **242**, 309–320 (1994).
6. E Anderson, et al., *LAPACK Users' Guide.* (Society for Industrial and Applied Mathematics, Philadelphia, PA), Third edition, (1999).

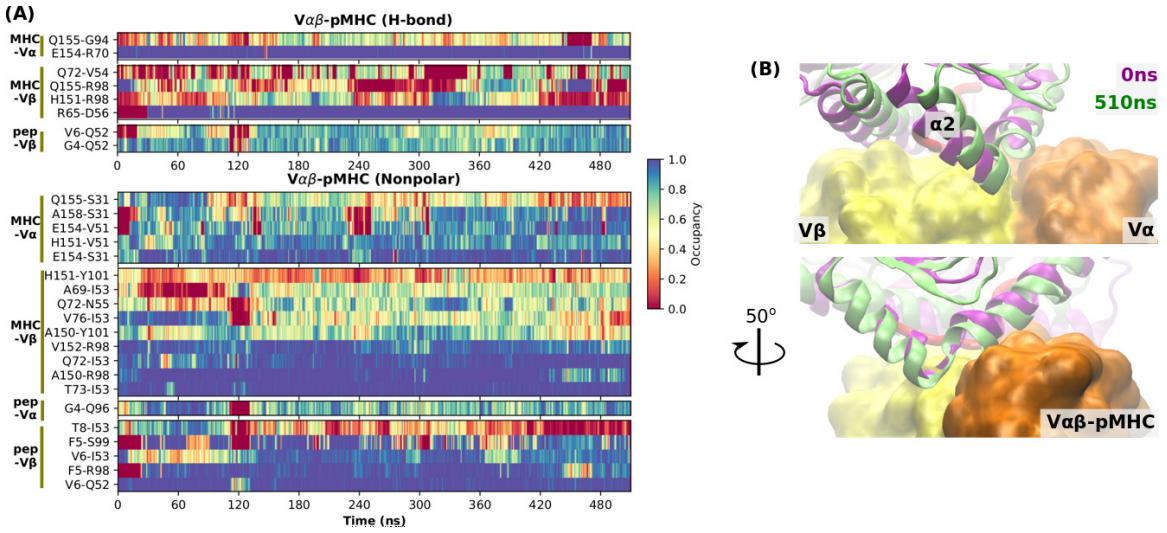


**Fig. S1.** The V$\alpha\beta$-pMHC interface. (A) Contact occupancy heat maps for V$\alpha\beta$-pMHC. The same occupancy cutoffs as in Fig. 2A,B were used. (B) Position of pMHC relative to the V-module. At the end of the simulation (510 ns), pMHC stays in a similar position as in the beginning of the simulation (*cf.*, Fig. 2D).

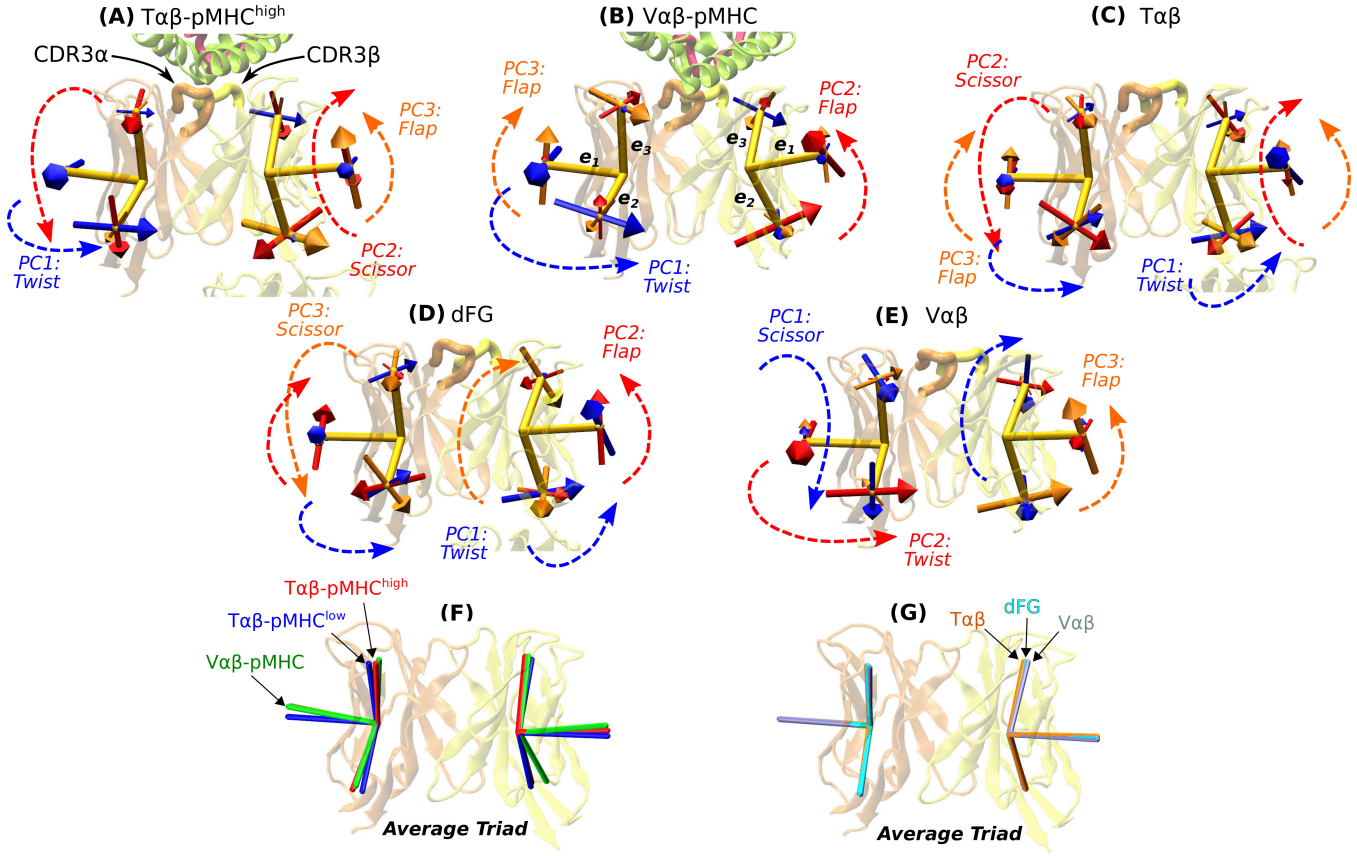


**Fig. S2.** V$\alpha$-V$\beta$ motion in different systems. (A–E) Directions of PC1–PC3. The same notations as in Fig. 3 are used. (A) T$\alpha\beta$-pMHC$^{high}$. (B) V$\alpha\beta$-pMHC. Both PC2 and PC3 appear to be a flap in different directions. (C) T$\alpha\beta$. (D) dFG. PC1 for V$\alpha$ is more clearly a twist, but for V$\beta$ it appears closer to flap (see also Movie S3). (E) V$\alpha\beta$. (F,G) Average triads for systems (F) with and (G) without a bound pMHC.

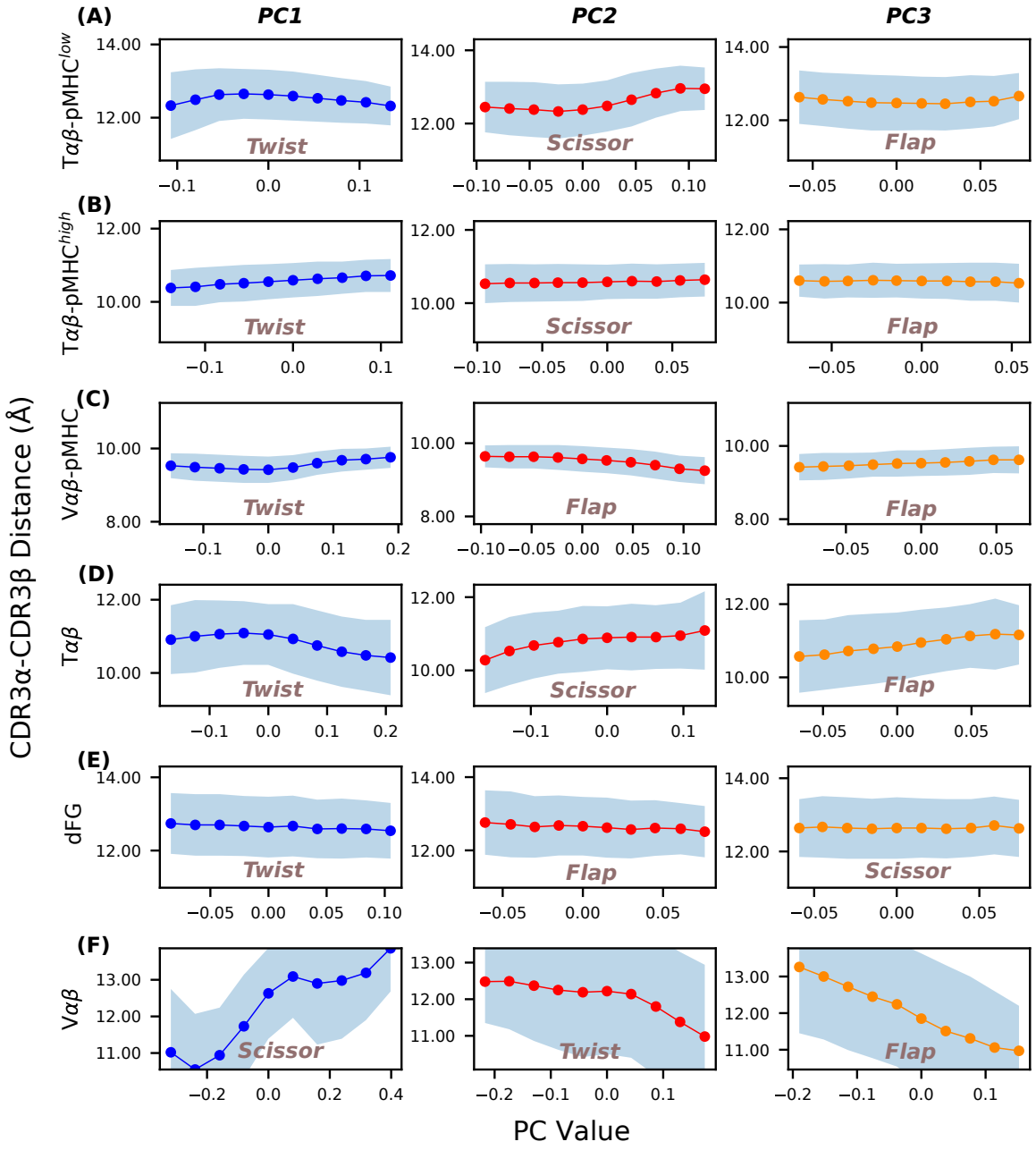


**Fig. S3.** Correlation between the CDR3 distance and PCs. Coordinate frames after the initial 200 ns were sorted by the values of the given PC mode (10 bins each). Within each bin, the CDR3 distance was averaged. Semi-transparent band: ±std. Vertical spans of all graphs were made the same, to aid with comparison. In (F), the least constrained V$\alpha\beta$ shows the largest variation in the CDR3 distance with PC values, which is consistent with the illustration in Fig. 3B. Also see Movie S2.

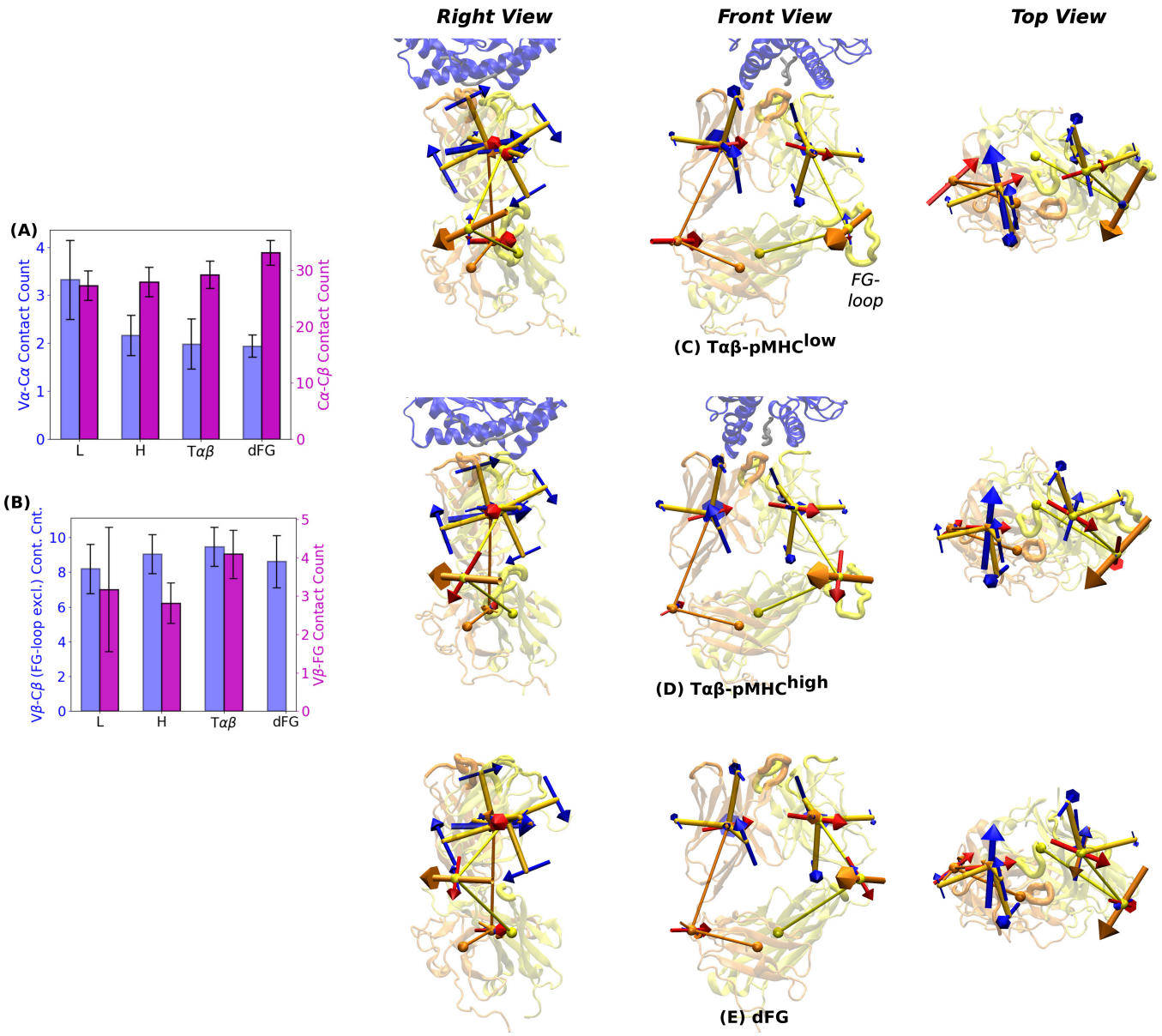


**Fig. S4.** Interdomain contacts and PCA of TCR$\alpha\beta$. (A,B) Contact count, using the same criteria as in Fig. 4A. 'L': T$\alpha\beta$-pMHC$^{low}$, 'H': T$\alpha\beta$-pMHC$^{high}$. (A) V$\alpha$-C$\alpha$ and C$\alpha$-C$\beta$. (B) V$\beta$-C$\beta$. Counts are separately with C$\beta$ excluding the FG-loop, and with the FG-loop only. (C–E) Directions of PCs for the BOC models of (C) T$\alpha\beta$-pMHC$^{low}$, (D) T$\alpha\beta$-pMHC$^{high}$, (E) dFG. The same notations as in Fig. 4B–D are used. Unlike PCs for the V-module (Fig. 3A,B, Fig. S2A–E), all systems show similar directions of PCs. Also see Movies S6,S7.

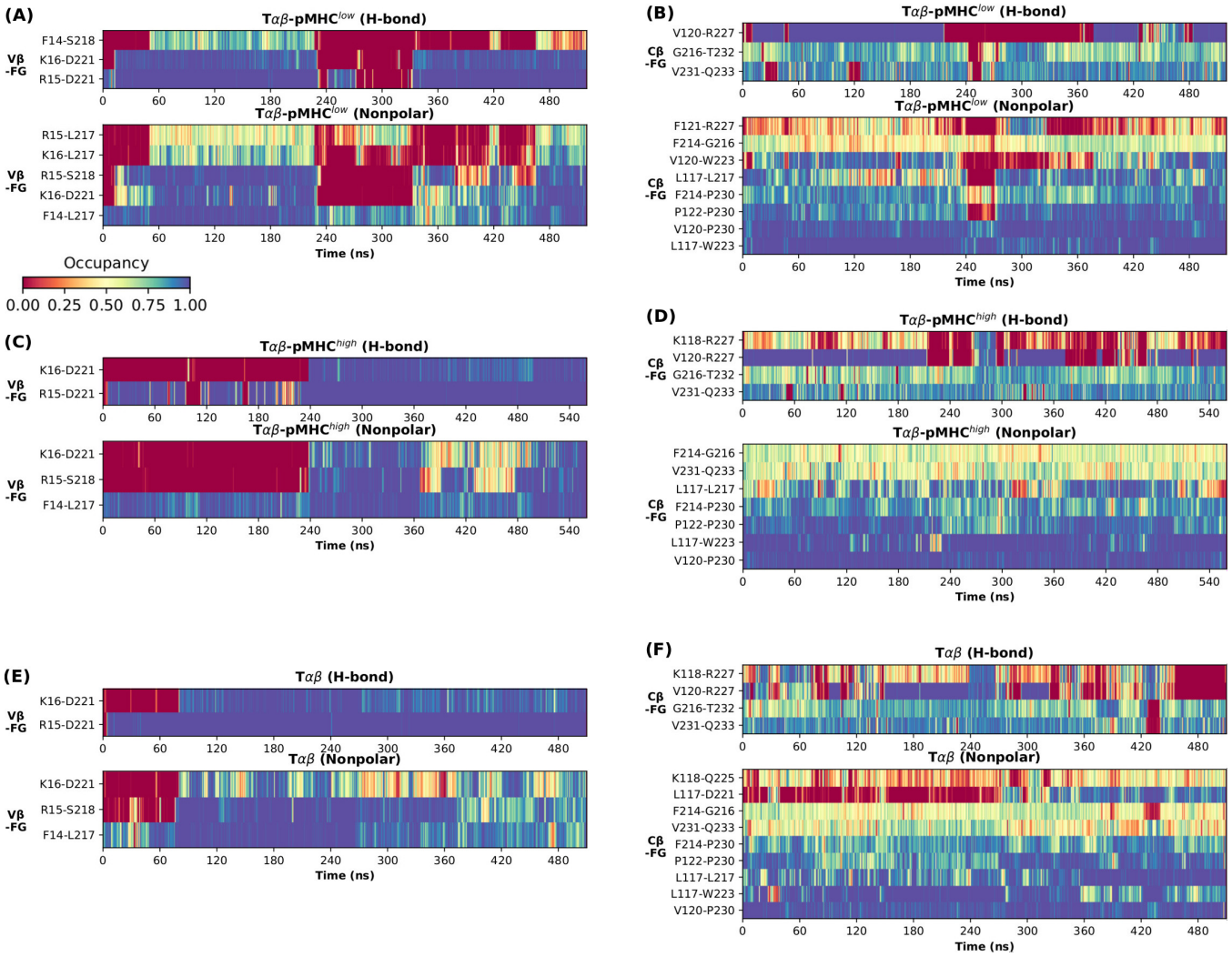


**Fig. S5.** Contacts within TCR$\beta$ involving FG-loop (G216–V231). (A,B) T$\alpha\beta$-pMHC$^{low}$, (C,D) T$\alpha\beta$-pMHC$^{high}$, (E,F) T$\alpha\beta$. (A,C,E) Contacts with V$\beta$. (B,D,F) Contacts with the rest of C$\beta$. All of these contacts are absent in dFG except for G216 and V231, which were kept in dFG in order to join the cleaved $\beta$-chain. They form contacts only with C$\beta$, and not with V$\beta$.

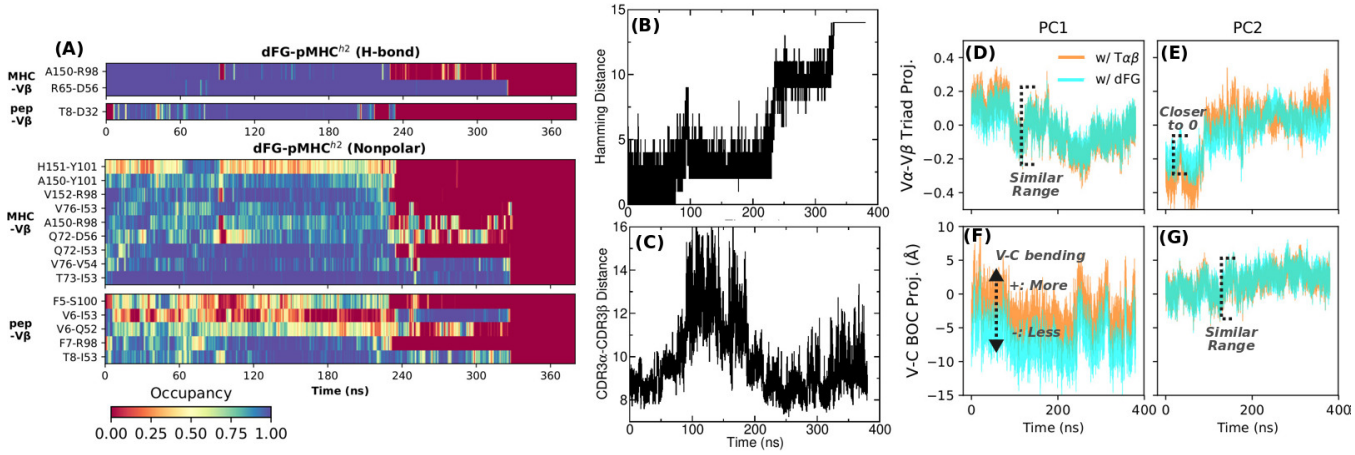


**Fig. S6.** Dissociation process of dFG-pMHC$^{h2}$. See Fig. 5 for explanation of panels. The complex completely dissociates at about 330 ns. Also see Movie S9.

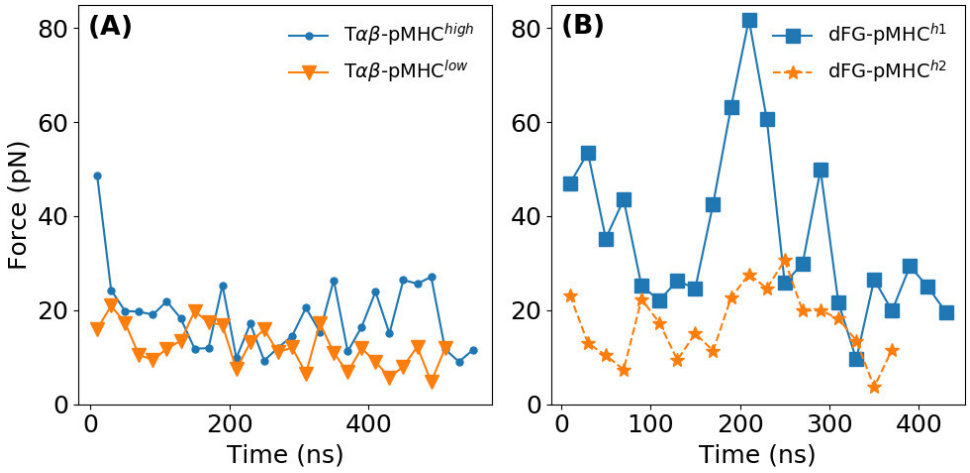


**Fig. S7.** Force on the complex measured with 20-ns time intervals. (A) T$\alpha\beta$-pMHC. The average force after 200 ns is 10.4 pN for T$\alpha\beta$-pMHC$^{low}$, and 16.4 pN for T$\alpha\beta$-pMHC$^{high}$. (B) dFG-pMHC.

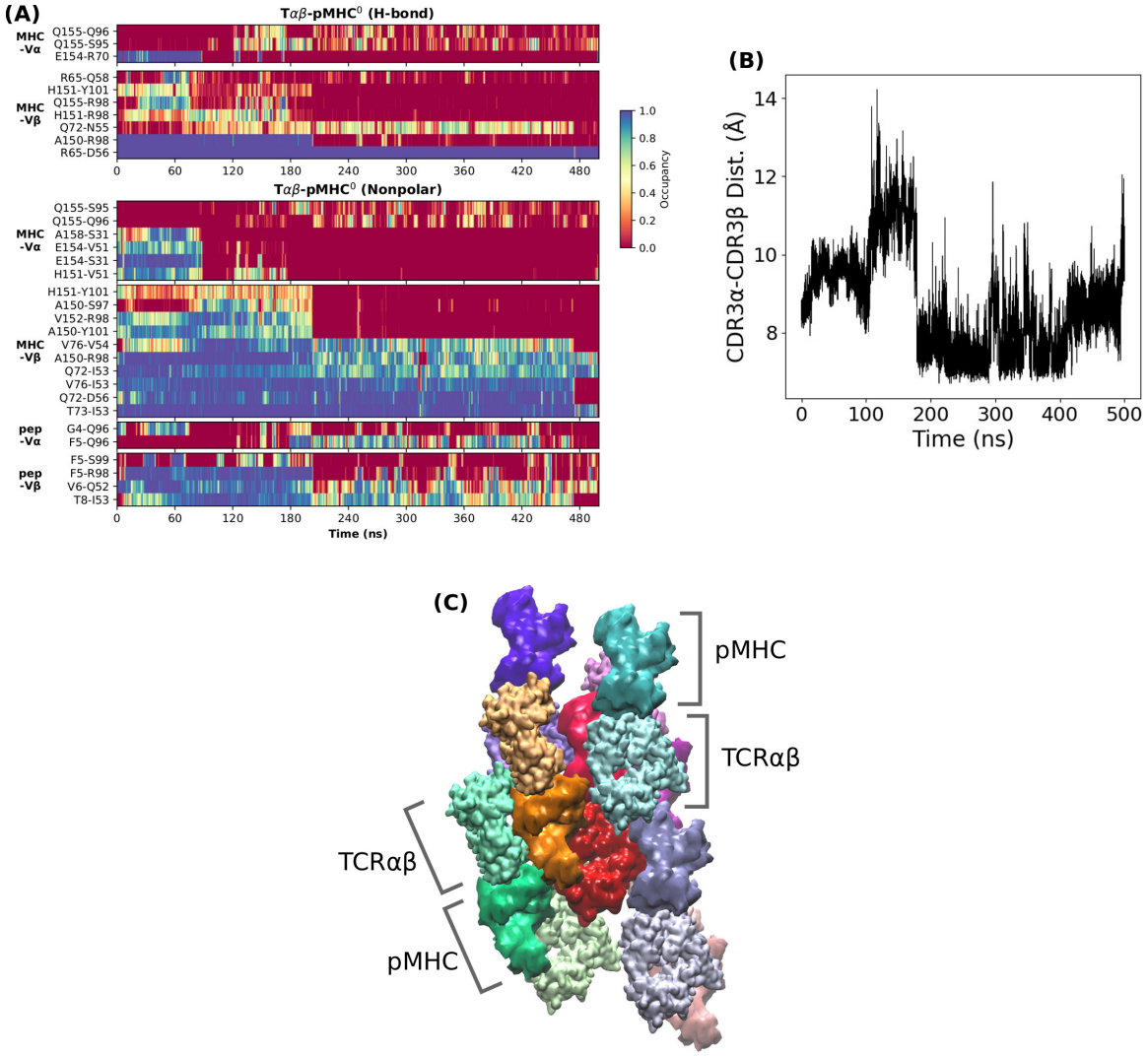


**Fig. S8.** Behavior of the unloaded case (T$\alpha\beta$-pMHC$^0$). (A) Contact occupancy trajectory. In order to monitor contacts that break early, a lower occupancy cutoff compared to Fig. 2A was used, with 10% average occupancy and 10% maximum instantaneous occupancy. By about 200 ns, nearly all pMHC-V$\alpha$ contacts as well as part of pMHC-V$\beta$ contacts break. (B) CDR3 distance trajectory. It increases and fluctuates before the major disruption of the interface. After the disruption, the CDR3 distance becomes low because the flexible CDR3$\alpha$ moves closer to CDR3$\beta$, attracted by the pMHC. But the distance still fluctuates widely compared to stable cases such as V$\alpha\beta$-pMHC and T$\alpha\beta$-pMHC$^{high}$ (Fig. 3E). (C) Crystal of PDB 1OGA, revealing longitudinal packing of the TCR$\alpha\beta$-pMHC molecule. In each complex, pMHC is rendered with a smoother surface. The TCR$\alpha\beta$-pMHC complex in the middle is colored red to illustrate the packed environment. For visualization, not all the surrounding molecules were built. Such packing likely suppresses conformational motion and helps with stabilizing the complex in the crystal.