



Cite this: *Phys. Chem. Chem. Phys.*,
2022, 24, 12397

Dynamics of peptide loading into major histocompatibility complex class I molecules chaperoned by TAPBPR†

Honglin Xu, Kaiyuan Song and Lin-Tai Da  *

Major histocompatibility complex class I (MHC-I) molecules display antigenic peptides on the cell surface for T cell receptor scanning, thereby activating the immune response. Peptide loading into MHC-I molecules is thus a critical step during the antigen presentation process. Chaperone TAP-binding protein related (TAPBPR) plays a critical role in promoting high-affinity peptide loading into MHC-I, by discriminating against the low-affinity ones. However, the complete peptide loading dynamics into TAPBPR-bound MHC-I is still elusive. Here, we constructed kinetic network models based on hundreds of short-time MD simulations with an aggregated simulation time of ~21.7 μs, and revealed, at atomic level, four key intermediate states of one antigenic peptide derived from melanoma-associated MART-1/Melan-A protein during its loading process into TAPBPR-bound MHC-I. We find that the TAPBPR binding at the MHC-I pocket-F can substantially reshape the distant pocket-B via allosteric regulations, which in turn promotes the following peptide N-terminal loading. Intriguingly, the partially loaded peptide could profoundly weaken the TAPBPR-MHC stability, promoting the dissociation of the TAPBPR scoop-loop (SL) region from the pocket-F to a more solvent-exposed conformation. Structural inspections further indicate that the peptide loading could remotely affect the SL binding site through both allosteric perturbations and direct contacts. In addition, another structural motif of TAPBPR, the jack hairpin region, was also found to participate in mediating the peptide editing. Our study sheds light on the detailed molecular mechanisms underlying the peptide loading process into TAPBPR-bound MHC-I and pinpoints the key structural factors responsible for dictating the peptide-loading dynamics.

Received 26th January 2022,
Accepted 11th May 2022

DOI: 10.1039/d2cp00423b

rsc.li/pccp

Introduction

Major histocompatibility complex class I (MHC-I) molecules, expressed by nucleated cells as a transmembrane protein, display endogenous peptide fragments on the cell surface for immune surveillance. These presented peptides are mainly produced by the ubiquitin-proteasome pathway, which involves the degradation of cytosolic proteins, such as regulatory proteins, misfolded and damaged proteins, virus-derived proteins in infected cells, and mutated proteins in cancer cells.¹ Then, the degraded peptides are transported into the endoplasmic reticulum (ER) through the transporter associated with antigen processing (TAP), which forms the base of the multisubunit membrane peptide-loading complex (PLC) in the ER. Meanwhile, newly synthesized empty MHC-I heterodimers can be recruited by calreticulin and become part of the PLC.²

Key Laboratory of Systems Biomedicine (Ministry of Education), Shanghai Center for Systems Biomedicine, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai 200240, China. E-mail: darlt@sjtu.edu.cn

† Electronic supplementary information (ESI) available. See DOI: <https://doi.org/10.1039/d2cp00423b>

Notably, peptide loading into MHC-I can be facilitated by one transmembrane chaperone protein tapasin which is also a key component of PLC. Tapasin preferentially catalyzes loading of the high-affinity peptides by blocking the low-affinity ones, known as “peptide proofreading” or “peptide editing”.^{3–7} Then, the optimal peptide-bound MHC-I (pMHC-I) complex will be released from PLC and transported to cell surface, thereby recognized by T cell receptors (TCR) on the cytotoxic T cells or natural killer (NK) cell receptors.⁸ Likewise, TAPBPR (TAP-binding protein related), serving as a second quality inspector, can also facilitate peptide loading into MHC-I independent of PLC and even accompany MHC-I beyond the medial Golgi.^{9–13} Besides, TAPBPR can aid the reglucosylation of suboptimally loaded or empty MHC-I, which is catalyzed by UDP-glucose:glycoprotein glucosyltransferase 1 (UGT1), resulting in the recycling of MHC-I to PLC (see Fig. S1, ESI†).¹⁴ Thus, TAPBPR serves as a crucial quality-control checkpoint during the antigen presentation pathway. In addition, soluble TAPBPR was also employed to load immunogenic peptides directly onto tumor cells rather than through the antigen presentation pathway inside the cell, implying a promising therapeutic application.¹⁵

A series of structural and functional studies regarding the TAPBPR-MHC-I complexes have shed light on the molecular mechanisms of the “peptide editing” process.^{10,13,16–22} Two recent crystal structures of the TAPBPR-MHC-I complex show that the concave N-domain of TAPBPR can encompass the MHC-I α 2-1 helix that flanks the pocket-F, resulting in a wedged helical conformation and widened pocket-F.^{16,17} In addition, one β -hairpin motif (known as jack hairpin) lies right beneath the floor of the peptide-binding groove and interacts with β_2m , suggesting its potential role in modulating the conformation of the peptide-binding groove.^{16,17} Importantly, one critical structural motif of TAPBPR, called “scoop loop” (SL), can directly insert into the pocket-F, thereby pre-occupying the C-terminal binding-site of incoming peptide.¹⁷ In addition, one TAPBPR residue, E105, could interact with the displaced MHC-I Y84 that otherwise forms direct contact with the C-terminus of the bound peptide. On the other hand, the C-terminal domain of TAPBPR could bind to the interface between the β_2m and α 3 domains of MHC-I.

Extensive experimental work has been devoted to investigating the functional roles of the SL region in regulating the peptide editing process.^{18,22,23} For example, by employing cell-based assays and site-directed mutagenesis analyses, Ilca *et al.* highlight the essential role of one SL residue L30 in mediating the peptide dissociation from MHC-I and find mutation of the SL region can significantly impair the “peptide editing” ability of TAPBPR.¹⁸ Moreover, former size-exclusion chromatography experiments indicate that the SL region contributes to stabilizing the empty MHC-I structures.²³ Their two-step peptide exchange assays further suggest that the SL region obstructs peptide rebinding and could be expelled by high-affinity peptides, therefore serving as a peptide filter.²³ The above experimental observations comply well with one crystal structure of the TAPBPR-MHC-I complexes in which the SL is found to occupy the pocket-F, thereby competing with the incoming peptides for the binding groove.¹⁷ Nevertheless, an alternative mode-of-action of TAPBPR was recently proposed, suggesting that the SL region might serve as a “peptide trap” to attract the high-affinity peptides by hovering over the pocket-F instead of inserting into the pocket.²² Despite the above experimental efforts, an atomic-level understanding of the functional role of TAPBPR in the peptide exchange process is still elusive due to the low spatiotemporal resolution.

Computational simulations allow us to investigate the conformational dynamics of biomolecules at atomic details and have been previously employed to study the MHC systems in the presence of peptide substrate or chaperones.^{19,22,24–27} Despite that, all former simulation work is limited to up to a few microseconds (μ s), which is still far less sufficient to capture large conformational rearrangements taking place at a relatively long timescale, *e.g.*, substrate binding or TAPBPR association, *etc.* Notably, one can now access the longer time-scale dynamics (*e.g.*, tens of μ s or even longer) by constructing Markov state models (MSMs) based on hundreds of short-time MD simulations.^{28–33} Recently, our and other groups have successively applied MSM to investigate the complete peptide loading dynamics into MHC-I molecules.^{34,35}

Here, to further examine the functional role of TAPBPR in the peptide loading process and the subsequent structural communications between peptide and TAPBPR, we constructed MSMs based on hundreds of 100 ns MD simulations with an aggregated simulation time of $\sim 21.7 \mu$ s. We uncover the key intermediate states of the antigenic peptide during its loading process in the presence of TAPBPR, which allows us to reveal the molecular mechanisms of how TAPBPR interactively interplays with the incoming peptides, thereby leading to its final dissociation from MHC-I. Moreover, we further dig deep into the structural couplings between pocket-F and remote pocket-B exerted by TAPBPR. Our work provides detailed molecular mechanisms underlying the TAPBPR-mediated peptide loading, which paves the way for the application of TAPBPR in future therapeutic use.

Results

Our former work indicates that the peptide N-terminal loading into MHC-I is the rate-limiting step during the substrate binding process when TAPBPR is absent.³⁵ Thus, to examine the regulatory role of TAPBPR in the peptide loading process, we here focus on investigating the loading dynamics of the peptide N-terminus into the pocket-B in MHC-I, with TAPBPR bound to the other end of the binding groove. To achieve this, we firstly constructed three TAPBPR-bound pMHC-I tertiary complexes (or TAPBPR-pMHC-I) using the crystal structure of TAPBPR (PDB id:5OPI)¹⁷ and three partially loaded pMHC-I models derived from our previous study (see Fig. 1A).³⁵ In specific, we selected three pMHC-I (ELAGIGILTV/HLA-A*0201) conformations from the identified intermediate state where the peptide N-terminus has already loaded into the MHC-I pocket-B but with the C-terminus still exposed in solvents (namely partially loaded). These pMHC-I conformations were chosen from each of the three identified free-energy basins (see Fig. S2, ESI†). Then, for each pMHC-I conformation, we modelled the TAPBPR-pMHC-I tertiary complex, and the final TAPBPR-pMHC-I models were subject to energy minimization (see Fig. 1B, see Materials and methods section for more details of the model constructions).

In these three TAPBPR-pMHC-I complexes, the peptide C-terminal (pT9-pV10) and connecting (pG4-pL8) regions (highlighted in green and orange in Fig. 1A and B, respectively) adopt distinct conformations, whereas the N-terminus (pE1-pA3) resembles each other (shown in cyan in Fig. 1A and B). In specific, the pE1 and pA3 backbones can form hydrogen bond (HB) interactions with the sidechain of MHC-I E63 and Y99, respectively, and the pE1 sidechain establishes salt-bridge interactions with MHC-I K66. In addition, the pL2 sidechain can pack into a hydrophobic pocket flanked by the MHC-I residues Y7, F9, A24, V34, M45, K66, and V67 (see Fig. 1C). On the other hand, the TAPBPR-SL region (residues K43-D56) occupies the binding groove of the peptide C-terminus *via* wedging into the MHC-I pocket-F. Particularly, the rim of SL (residues L51-S54) hovers over the binding groove and interacts with the α 1 helix of the MHC-I HC *via* polar contacts (see Fig. 1D).

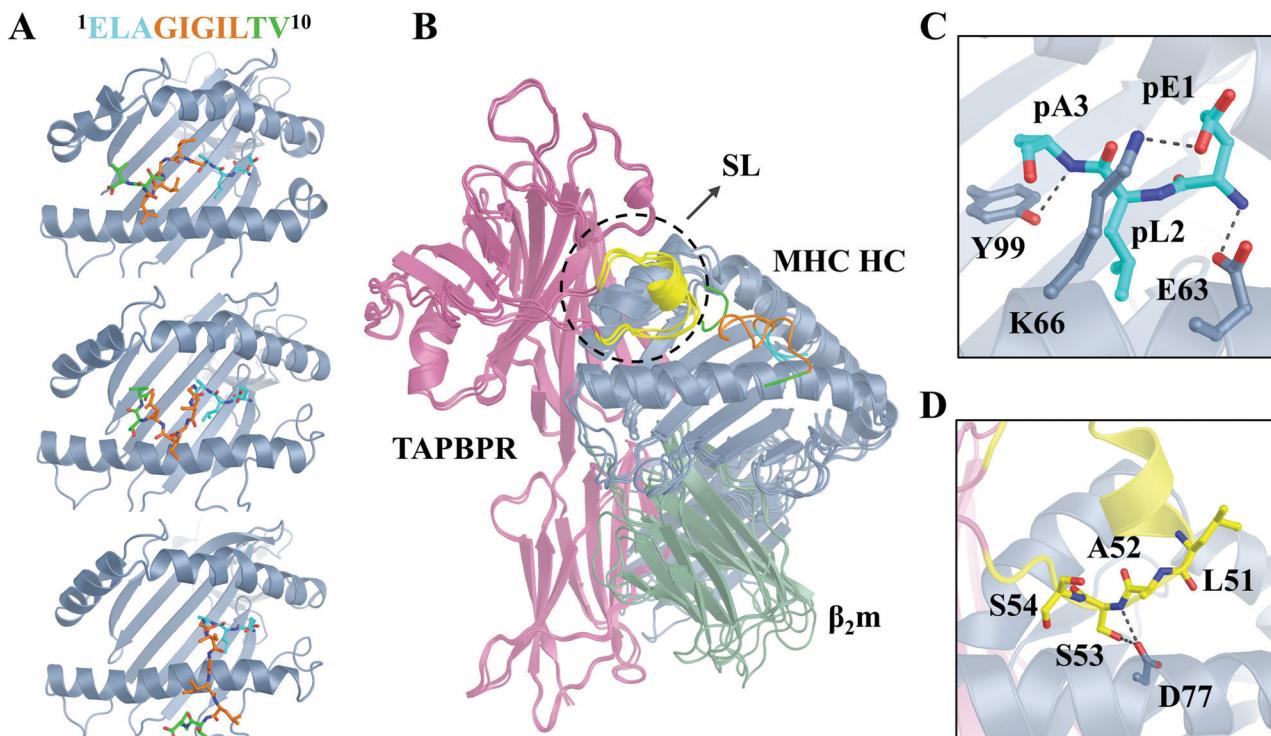


Fig. 1 Construction of TAPBPR-pMHC-I tertiary complex and energy minimization. (A) Three constructed pMHC-I complexes where the antigenic peptides are partially loaded but adopt distinct conformations. The MHC-I molecule is in blue; the peptide N-terminus (E1-A3), connecting region (G4-L8), and C-terminus (T9-V10) is shown in cyan, orange, and green sticks, respectively. (B) Constructed three TAPBPR-pMHC-I tertiary complexes based on the above pMHC-I conformations, with the TAPBPR shown in purple cartoon except for the SL region that is highlighted in yellow. The MHC heavy chain (MHC HC) and β_2 -microglobulin (β_2 m) is a blue and green cartoon, respectively. (C) Interaction network between peptide N-terminus and MHC-I residues. (D) Highlight of the binding interface between the TAPBPR-SL region and MHC-I residues. In C and D, critical hydrogen bonds (HBs) are indicated with black dashed lines.

Based on each of the above constructed TAPBPR-pMHC-I complexes, we firstly derived several peptide dissociation trajectories using steered MD (SMD) simulations, which presumably serve as the reverse process of the peptide loading pathway. Next, we selected 100 representative conformations from the above SMD simulations, based on which we performed extensive unbiased 100 ns MD simulations (see Materials and methods section for more details of the setups of the SMD and unbiased MD simulations). Finally, we collected a simulation dataset with an aggregated simulation time of $\sim 21.7 \mu\text{s}$ for the final MSM construction (see Materials and methods section for the details of the MSM construction and validation).

MSM unveils four metastable states during the peptide loading process into TAPBPR-MHC-I

With TAPBPR bound in the pocket-F, our MSM captures four metastable states during the peptide N-terminal loading into MHC-I, namely S1-S4 (see Fig. 2A). S1 represents the unbound state where the peptide N-terminus is largely exposed in the solvents, whereas S2-S4 correspond to the partially bound state. Notably, although S2 and S4 seem indistinguishable from the conformational projection onto the top two time-structure based independent component (tIC1, tIC2), we can observe a clear state-separation between S2 and S4 along tIC3 (see Fig. S3, ESI[†]). Further transition path theory (TPT) analyses determine the

dominant N-terminal peptide loading path, following the transition path of S1 \rightarrow S2 \rightarrow S3 (see Fig. 2B). Notably, insertion of the peptide N-terminus into the pocket-B can be accomplished in the S1 \rightarrow S2 transition, reflected from the striking decrease of the RMSD value of the peptide N-terminus from $\sim 9 \text{ \AA}$ to $\sim 1 \text{ \AA}$, which maintains stable in the following S2 \rightarrow S3 transition (see Fig. 2C). On the other hand, the TAPBPR-SL region does not demonstrate considerable structural differences between the four states based on the RMSD calculations. Nevertheless, one structural motif of SL, L51-S54 (or SL-LASS), undergoes significant conformational deviations, particularly in the S2 \rightarrow S3 transition, with the calculated RMSD value raised from $\sim 2.4 \text{ \AA}$ to 4 \AA (see Fig. 2C). Alternatively, the S2 state can also transit to an off-path state S4 in which both peptide N-terminus and the SL-LASS motif display larger structural flexibilities compared to the conformations in S2. The TPT analyses also indicate that the net flux of the S2 \rightarrow S3 transition (1.86×10^{-4}) is considerably higher than that for the S2 \rightarrow S4 transition (5.17×10^{-6}), proving that the former transition is more likely to happen.

Specifically, in S2 and S3, the peptide pE1-pA3 motif can establish quite stable HBs with the sidechains of MHC-I residues E63, K66, Y99, and Y159 via either the sidechain (*i.e.*, E1) or backbones (*i.e.*, E1 and A3) (denoted as HB1-HB4 in Fig. 3A and B). In S4, the HB between A3 and Y99 is broken,

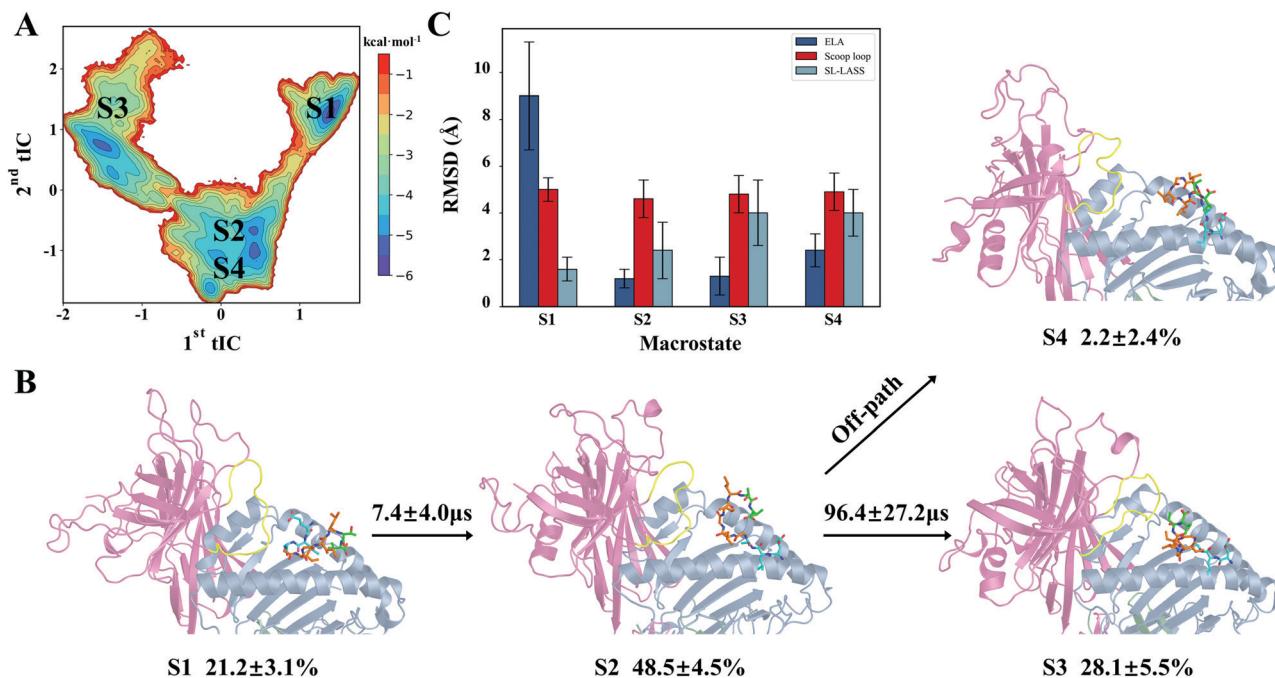


Fig. 2 The MSM reveals four macrostates during the peptide loading into MHC-I in the presence of TAPBPR. (A) Free energy projection of the MD conformations onto the top two tICs. The location of each macrostate is labeled. (B) Representative conformation of each macrostate that was randomly selected from the most populated microstate for a given macrostate. The equilibrium population of each state and the MFPTs for each state-to-state transition are also provided. (C) Root mean square deviation (RMSD) of the peptide N-terminus (ELA), scoop loop of TAPBPR, and the SL-LASS motif for each macrostate, calculated for the corresponding C_x atoms with respect to the initial conformation.

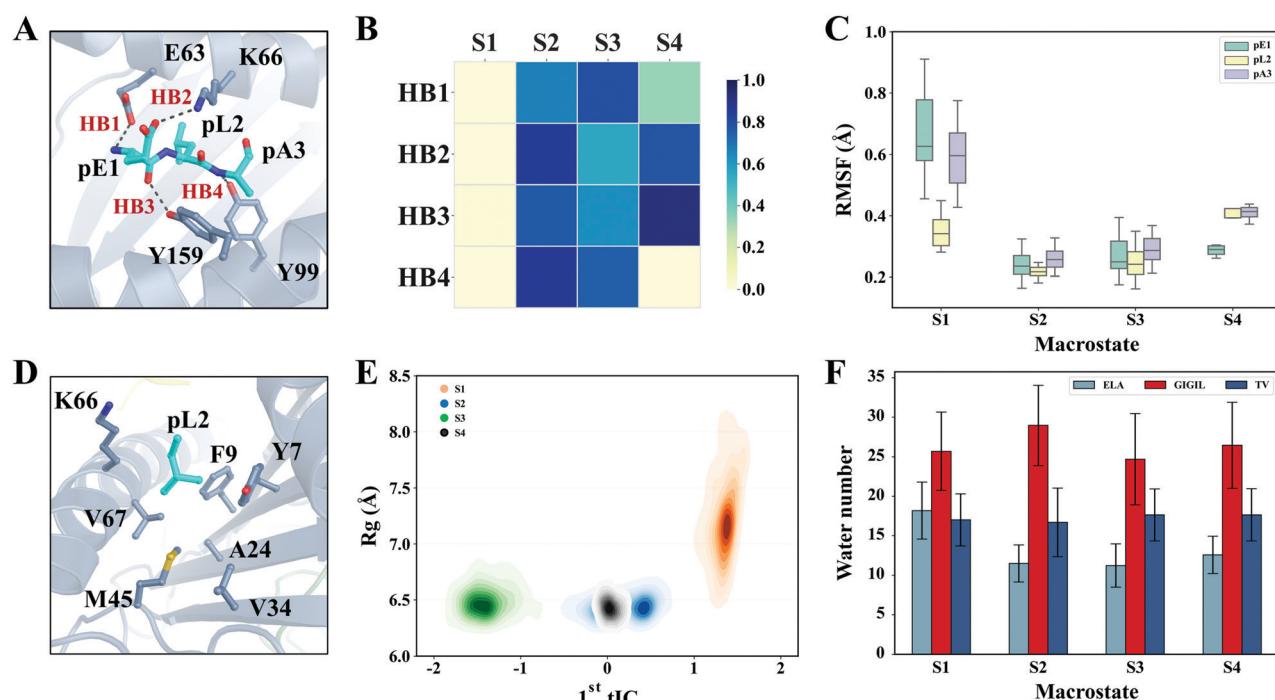


Fig. 3 Structural analyses of the identified macrostates. (A) Highlights of four HBs between peptide N-terminus and MHC-I residues (HB1-HB4). (B) The propensity of the HBs shown in (A) for each macrostate. (C) Root mean square fluctuation (RMSF) of the C_x of three peptide N-terminal residues for each macrostate. (D) Structural illustration of the hydrophobic contacts between pL2 and pocket-B residues after N-terminal insertion. (E) Projection of MD simulations onto two coordinates: the first tIC and the radius of gyration (R_g) of the residues shown in (D). (F) The number of water molecules surrounding each of three peptide segments, i.e., N-terminus (ELA), connecting region (GIGIL), and C-terminus (TV), with a distance cutoff of 3 Å.

which leads to profound increased fluctuations of pL2-pA3 (see Fig. 3C). This result complies with the above RMSD calculations of peptide N-terminus in S4 (see Fig. 2C). Considering the fact that TAPBPR can discriminate the high-affinity peptides from the low-affinity ones, S4 thus might be a biologically relevant state for the low-affinity peptides of MHC. Moreover, pL2 can accommodate into a hydrophobic pocket-B flanked by MHC-I residues Y7, F9, A24, V34, M45, K66, and V67 (see Fig. 3D). These non-polar contacts can be formed immediately after the peptide N-terminus penetrates into pocket-B and remain stable in the following transitions, as demonstrated by the radius of gyration (R_g) calculations (see Fig. 3E). Notably, the N-terminal loading is accompanied by a significant loss of the surrounded water molecules by ~ 7 (see Fig. 3F). In contrast, the C-terminal and the connecting regions are largely exposed in the solvents in all states.

Finally, to examine whether the 4-state model is sufficient to describe the peptide-loading mechanism, we also lumped the 500 microstates into five macrostates. The resulting 5-state model shows that in addition to the four metastable states identified above, one more state, S5, can be isolated from the S2 state of the 4-state model, and well-separated from other states along tIC3 (see Fig. S4A and B, ESI[†]). Further structural

analyses indicate that the TAPBPR-SL region in S5 contains a helical motif, which is similar to the SL conformation of the crystal structure (see Fig. S4C and D, ESI[†]). Despite that, the peptide ELA motif as well as the SL region remain stable, resembling the structural features of S2 (see Fig. S4E, ESI[†]). We thus conclude that the 4-state model is sufficient to describe the complete peptide-loading event.

The SL region is structurally highly heterogenous

The SL region was proposed to be highly flexible as inferred from former experimental studies.²² We thus further examined the structural heterogeneity of the SL region and evaluated to what extent our MD simulation might deviate from the reported crystal structure (PDB id: 5OPI).¹⁷ To achieve this, we first calculated the backbone RMSD of the SL residues K43-D56 for all MD conformations relative to the crystal structure by fitting to the MHC helices (see Fig. 1B). The results clearly show that compared to the crystal form, the SL conformations observed from our MD simulations are structurally highly heterogeneous, with the calculated RMSD ranging largely from 2–7 Å, for S3 particularly, the structural difference can reach up to ~ 10 Å (see Fig. 4A). In addition, given the fact that the SL G45-A52 motif displays a helical conformation in the crystal

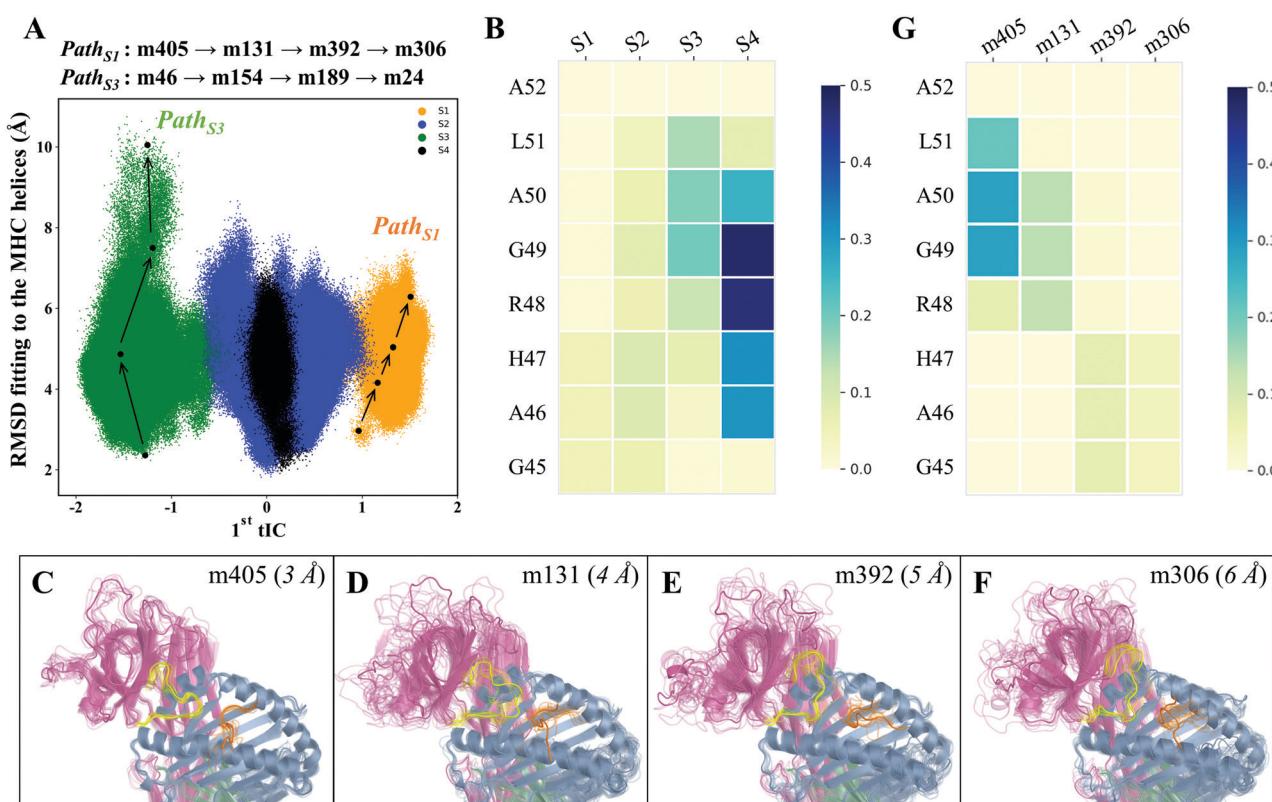


Fig. 4 The SL region displays heterogenous structural features from our MD simulations. (A) Scatter plot for all MD conformations projected onto two reaction coordinates: the first tIC; the backbone RMSD of the SL residues K43-D56 referenced to the crystal conformation (shown in Fig. 1B) after fitting to the MHC helices (Q54-Y84 & M138-Y171). S1–S4 are illustrated in orange, blue, green, and black, respectively. Two representative SL-opening pathways in S1 and S3, namely *Path_{S1}* and *Path_{S3}*, were determined using TPT based on our 500-state MSM, with the detailed transitions shown above. (B) Helical propensity of the SL G45-A52 motif for S1–S4. (C–F) Representative conformations of each microstate involved in *Path_{S1}*, demonstrating a respective RMSD value of ~ 3 Å (C), 4 Å (D), 5 Å (E), and 6 Å (F) for the SL residues K43–D56 referenced to the crystal form. For each microstate, ten randomly selected MD conformations were illustrated. (G) Helical propensity of the SL G45-A52 motif for each microstate involved in *Path_{S1}*.

structure, we then further calculated the helical propensity of this motif for S1–S4. The results indicate that S1 and S2 exhibit lower helical propensities compared to S3 and S4 (see Fig. 4B). Notably, the off-state S4 contains the highest helical conformation among all states, indicating that S4 likely resembles the crystal structure the most. Indeed, the calculated RMSDs after fitting to the crystal SL itself indicate a more crystal-like conformation for S4 compared to other states (see Fig. S5A and B, ESI[†]). Taken together, our MD simulations indeed explored diverse structural conformations of SL during the peptide-loading process.

Next, to closely inspect the SL structure, we zoomed into two representative structural rearrangements of the SL region involved in S1 and S3 that represents the states before and after peptide binding, respectively. In specific, we employed the TPT analyses to search, at microstate level, for the dominant transition path of SL from a low-RMSD to a high-RMSD conformation (namely *Path*_{S1} and *Path*_{S3} in Fig. 4A). We found in *Path*_{S1}, *via* the microstate-level transitions m405 → m131 → m392 → m306, the m405 and m131 states locate closer to the MHC-I pocket-F than m392 and m306, with an RMSD difference of ~3 Å and ~4 Å to the crystal form, respectively (see Fig. 4C–F). Meanwhile, the R48-L51 motif displays relatively higher helical propensity in m405 and m131, suggesting that approaching the pocket-F facilitates the formation of secondary structure, due likely to the rigidified conformation conferred by MHC. In comparison, transiting to m392 & m306 leads to the loss of the secondary structure of the R48-L51 motif, caused largely by its increased flexibilities after deviating from the pocket-F (see Fig. 4G). A similar tendency can also be found for *Path*_{S3} where the

low-RMSD states, m46 and m154, exhibit higher helical conformation than the high-RMSD states (see Fig. S6A–E, ESI[†]). Particularly in m24, displaying a RMSD difference ~10 Å to the crystal form, the SL region is mostly out of the pocket-F (see Fig. S6D, ESI[†]).

The peptide N-terminal loading into pocket-B destabilizes the TAPBPR-SL region

The biological function of TAPBPR is considered to facilitate the binding of high-affinity peptides by preventing low-affinity ones.^{10,13,14} We therefore further examined the structural features of the TAPBPR-SL region upon peptide N-terminal loading and how the SL might interplay with the incoming peptide. We found that before peptide binding (*i.e.*, in S1), the SL-S53 could form three HBs with two MHC-I helices *via* K146 and D77 (denoted as HB1–HB3 in Fig. 5A and B), which contributes to stabilizing the SL region in pocket-F. Intriguingly, the loading of peptide N-terminus into pocket-B can profoundly weaken the above HB networks (in S2 and S3), particularly in S3 where only the HB between the S53 sidechain and D77 was preserved (see Fig. 5A and B). Transition to S4 results in impairing all the above HBs. In addition, we further calculated the C_α–C_α distances between TAPBPR-SL S53 and MHC-I D77 & K146 (denoted as d1 and d2 in Fig. 5C, respectively). We found a considerable increase of d2 and a large fluctuation of d1 in S3 compared to other states (see Fig. 5C). Further *R*_g calculations for TAPBPR-SL S53 and MHC-I residues D77, L81, Y116, Y123, T143, and W147 also indicate a tendency of SL dissociation from pocket-F in S3, reflected from a larger *R*_g value compared to

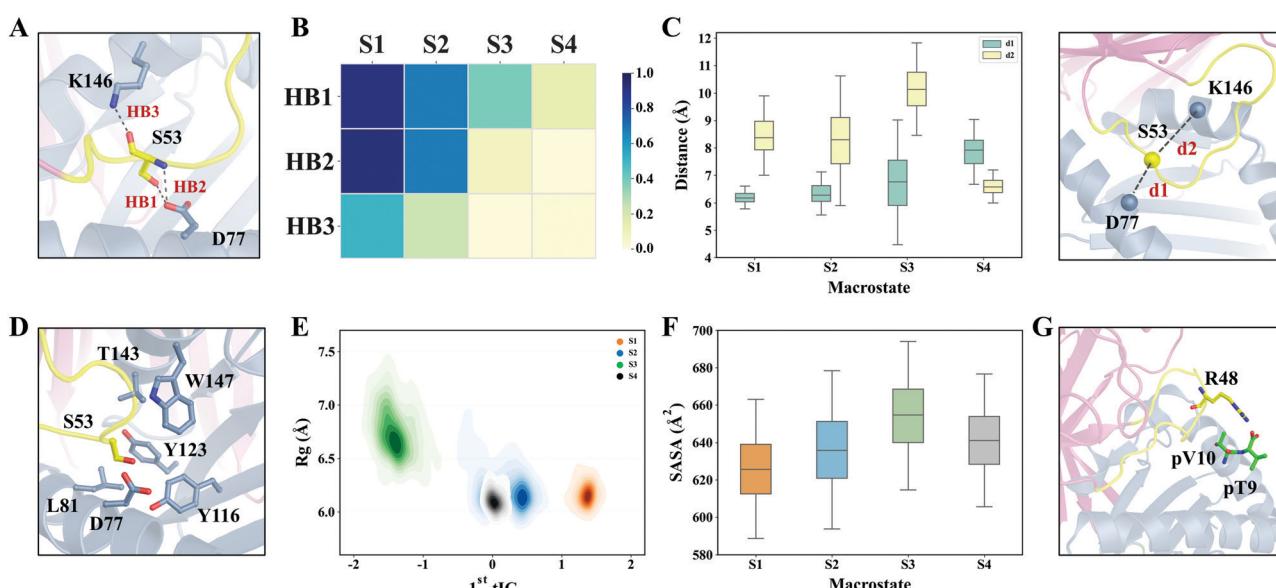


Fig. 5 Structural analyses of the SL-MHC interacting interface during the peptide loading process. (A) Highlights of three HBs between TAPBPR-SL and MHC-I pocket-F residues (HB1–HB3). (B) The propensity of the HBs shown in (A) for each macrostate. (C) The C_α–C_α distance between two pairs of residues: SL-S53 and MHC-K146; SL-S53 and MHC-D77. The structural illustration is also provided on the right panel. (D) Structural highlight of the pocket-F residues and SL-S53. (E) Projection of MD simulations onto two coordinates: the first tIC and the radius of gyration (*R*_g) of the residues shown in (D). (F) Solvent accessible surface area (SASA) of the SL-LASS motif for each macrostate. (G) The structural highlight of the salt-bridge interaction between SL-R48 and the peptide C-terminus in S3. In (C) and (F), the mean values were calculated by averaging over all the microstates that belong to one certain macrostate, the corresponding errors were then calculated.

other states (see Fig. 5D and E). The above results imply that the SL-LASS motif tends to leave the pocket-F in the S2 → S3 transition, which is further supported by the largest solvent-accessible surface area (SASA) value of the SL-LASS motif in S3 (see Fig. 5F). Interestingly, we also captured a direct contact between the SL-R48 and the peptide -COO⁻ group in S3, accounting for ~36% of the S3 conformations (see Fig. 5G). Taken together, the peptide N-terminus loading into pocket-B can impose considerable impacts on the stability of the SL region, *via* both direct contacts and allosteric effects, which promotes the TAPBPR dissociation from MHC-I.

In addition to the significant structural changes of TAPBPR-SL in the S2 → S3 transition, we also observed a significant unfolding of the MHC-I α 2-1 helix (residues Q141-K144) that is adjacent to the pocket-F (see Fig. 6A). In S1 and S2, the α 2-1 region maintains a fairly well helical form, only the terminal residues Q141-T142 display relatively low helical propensity (see Fig. 6B). In contrast, the S2 → S3 transition can lead to a great decrease of the helical contents for T142, T143, and K144 (all <20%) (see Fig. 6B). The above helical unfolding results in large structural orientations of T142 and T143. In specific, T142 can establish two stable HBs with E123 and N145 from the TAPBPR N-domain in S3 (see the left panel in Fig. 6C and D); T143, originally forming HB interaction with MHC-Y123 in S1 and S2, loses this contact after transiting to S3, caused likely by the buckling effect exerted by the rotation of the adjacent MHC-W147 (see the right panel in Fig. 6C and D, and Fig. S7A and B, ESI[†]).

Based on the transition count matrix, we found a total of 262 transitions from S2 to S3, assigned to eight different MD trajectories. In particular, we selected one representative S2 → S3 transition, belonging to the microstate-level transition of micro22(S2) → micro472(S3). The SL region in micro22 tends to be closer to pocket-F compared with that in micro472, as implied from the R_g calculations (see Fig. S8A and B, ESI[†]). Moreover, the SASA of the SL-LASS motif is also smaller in micro22 (see Fig. S8C, ESI[†]). These results comply well with the

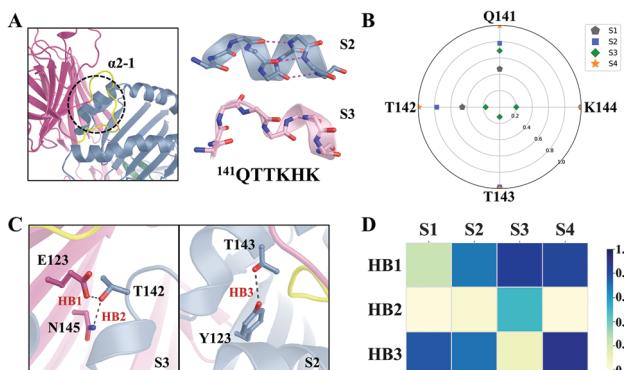


Fig. 6 Unfolding of the MHC α 2-1 helix during the S2 → S3 transition. (A) The MHC α 2-1 motif (Q141-K146) positions near to SL, encircled on the left panel. Highlights of the α 2-1 conformations in S2 and S3 are provided on the right panel. (B) Helical propensities of four α 2-1 residues (Q141-K144) in each metastable state. (C) Highlights of three HBs between the α 2-1 motif and TAPBPR (left) and MHC-I pocket-F residues (right) (HB1-HB3). (D) The propensity of the HBs shown in (C) for each macrostate.

above conclusion that the S2 → S3 transition can potentially drive the dissociation of SL from pocket-F. In addition, our Mean First Passage Time (MFPT) calculations indicate that the S2 → S3 transition is the rate-limiting step during the peptide loading process and is one order of magnitude slower than the S1 → S2 transition. The rationale behind this is likely due to the above observations that the peptide N-terminal loading in the S1 → S2 transition imposes inconsiderable structural perturbation on the TAPBPR-SL region, whereas significant conformational changes of SL were observed during the S2 → S3 transition (see Fig. 5A-F). Moreover, the unfolding of the MHC-I α 2-1 helix highlighted above also contributes to slowing down the S2 → S3 transition rate.

The TAPBPR binding in pocket-F regulates the peptide-binding groove allosterically

To further investigate the dynamic communications between pockets-B and F, we performed dynamical cross-correlation matrix (DCCM) analysis for MHC-I in different contexts, including apo MHC-I (peptide free), pMHC-I, and TAPBPR-pMHC-I (partially loaded). To achieve this, we employed our formerly conducted MD simulations for the apo MHC-I and pMHC-I,³⁵ each with three parallel 100 ns MD simulations (see Materials and methods section for the details of the simulation setups). For TAPBPR-pMHC-I, we chose MD trajectories that completely belong to S1, S2, and S3 from our original simulation dataset. The DCCM

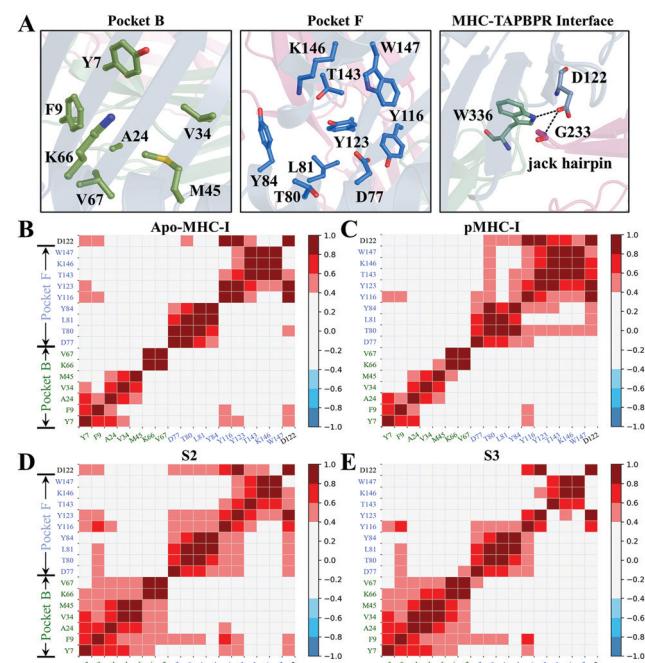


Fig. 7 Cross-correlation analyses for MHC-I in varied contexts. (A) Structural highlights of the residues used for correlation analyses in pocket-B (left), pocket-F (middle), and MHC-TAPBPR interface (for the jack hairpin region, right panel). The dynamical cross-correlation matrices were calculated for apo MHC-I (B), pMHC-I binary complex (C), S2 (D), and S3 (E). The correlation coefficients are averaged over the MD simulations that only belong to a given macrostate, with the coefficient above 0 indicating a positive correlation, below 0 indicating a negative correlation.

analysis was calculated for Y7, F9, A24, V34, M45, K66, V67 in pocket B and D77, T80, L81, Y84, Y116, Y123, T143, K146, W147 in pocket F (see the left and middle panels in Fig. 7A).

The results clearly show that the presence of peptides can profoundly enhance the mutual communications inside pocket-F, particularly for T80 and Y84, although that in pocket-B is barely altered (compare Fig. 7B and C). In sharp contrast, in S1, S2 and S3, the structural correlations within both pocket-B and F are greatly strengthened compared to that in both apo-MHC-I and pMHC-I (see Fig. 7B-E and Fig. S9A, ESI[†]). Notably, in S2, the presence of TAPBPR leads to strong couplings between pocket-B and F, mainly *via* MHC-I F9, Y116, and Y123. In comparison, noticeable reduced correlations between pocket-B and F can be observed in S3 (see Fig. 7E). In particular, Y123 in pocket-F decorrelates with Y7 and F9 in pocket-B after the S2 → S3 transition, which is likely caused by the loss of the T143-Y123 contact in S3 induced by the unfolding of the α 2-1 helix, as highlighted above (see Fig. 6C, D and 7E). In addition, Y123 also decouples with most residues within the pocket-F. Moreover, we also observed impaired couplings between MHC-I D122 and pockets B & F in S3 (see Fig. 7E), likely due to the rotation of D122 from the peptide-binding groove. Simultaneously, the HB formed between D122 and TAPBPR jack-hairpin residue G233 in S2 disappears after transitioning to S3 (see the right panel in Fig. 7A and Fig. S9B, ESI[†]). Despite that, D122 maintains its HB interaction with MHC-I W336 in all states (see Fig. S9B, ESI[†]).

We further investigated the detailed allosteric paths between pocket B and F in MHC-I for S1–S3 by employing the optimal correlation network analyses.^{36,37} In specific, we selected the C_α atoms of V34 and Y123 as the reference nodes for pocket B and F, respectively, due to their important roles in stabilizing the incoming peptides in bound state. We then determined the dominant correlation path connecting the two reference nodes. Our results show that the identified allosteric paths for S1–S3 could all propagate *via* two common nodes, namely R97 and Y116, indicating their critical roles in communicating two distant pockets in MHC (see Fig. S10, ESI[†]). Moreover, S1 and S3 share exact the same optimal path, and comparing to S1, the correlations between V25 and Y116 is relatively stronger in S3. This finding is consistent with the observation that the partially-loaded peptide in pocket-B could tighten the coupling between pocket-B and pocket-F. Taken together, the peptide N-terminal loading into pocket-B with TAPBPR bound in pocket-F not only strengthens the interaction networks within each pocket but also allosterically regulates the cross-talk between two pockets.

Discussion

The presence of TAPBPR tightens the couplings between pocket-B and F

Our former work has revealed the complete peptide-loading dynamics into MHC-I without TAPBPR, and found that insertion of the peptide N-terminus into pocket-B limits the overall

loading rate. The following C-terminal binding, however, is a relatively fast process.³⁵ The current study extends the understanding of how the presence of TAPBPR might exert an influence on the peptide-loading dynamics. Our DCCM analyses indicate much stronger couplings within pocket-B in TAPBPR-bound states (S1, S2, and S3) compared to the apo MHC-I (see Fig. 7B-E and Fig. S9A, ESI[†]). In addition, obvious dynamic correlations between pocket-B and F were also observed in both S2 and S3 where the peptide N-terminus is in a bound state but the C-terminus is unbound. Thus, our work suggests that TAPBPR binding likely stabilizes MHC-I in a peptide-receptive state, which promotes the N-terminal loading of the incoming peptide. This result is consistent with our previous observation that TAPBPR can re-position the critical pocket-B residues in ideal orientations ready for recognizing the incoming peptide.³⁵

Additionally, our findings also comply well with former experimental investigations performed using biochemical, X-ray, and NMR techniques.^{10,13,18,19,23,38} For example, previous mutagenesis studies suggest that mutations of the SL residues can impair the peptide-editing function of TAPBPR, which, according to our results, is likely caused by the loss of the functional role of TAPBPR in modulating the pocket-B.^{18,23} Furthermore, the crystallographic study of tapasin demonstrates that H2-K^b bound with dipeptide GL (one SL motif of tapasin) displays very similar conformation of the binding groove as the fully peptide-loaded state, suggesting the tapasin's role in modulating the substrate pocket prior to peptide binding.³⁸ Solution NMR studies have also observed allosteric communications between pocket-A and -F in MHC catalysed by TAPBPR.^{19,21} Moreover, MHC-I residues F9, R97, Y116, and Y123 are found to be responsible for modulating allosteric communications between two pockets. Former mutagenesis work has also indicated that the Y116D substitution in MHC-I can greatly weaken the TAPBPR binding with MHC-I or impair the peptide exchange ability of TAPBPR.^{18,20} Combined with the fact that Y116 is highly conserved across A2 and A24 superfamilies that are strong TAPBPR binders,¹⁸ it can be expected that the Y116 might serve as a critical site to collaborate pocket-F and B upon TAPBPR binding.

The binding of peptide N-terminus in pocket-B promotes the TAPBPR-SL dissociation from pocket F

TAPBPR is proposed to act as a “peptide editor” that promotes the binding of high-affinity peptides but expels the low-affinity ones.^{10,13} However, how exactly TAPBPR interplays with the substrate peptide after its N-terminal loading into pocket-B is still elusive. Our current work suggests that the peptide binding can profoundly destabilize the TAPBPR-SL region, *via* both allosteric modulation or direct competition with the peptide C-terminus. In particular, the peptide loading can substantially promote the dissociation of the SL-LASS motif from the pocket-F, accompanied by the unfolding of the adjacent MHC-I α 2-1 helix region. Notably, the peptide C-terminus can also establish direct contacts with TAPBPR R48 in S3, which likely further drives the SL dissociation. Consistent with our findings, the former crystal study suggests that the peptide C-terminus engages in dislodging

the part of the tapasin-SL after binding of the peptide N-terminus in pocket-B.³⁸ Likewise, a recent fluorescence polarization study for TAPBPR-SL variants also indicates that SL serves as a peptide selectivity filter either through direct steric clashes with C-terminus or allosteric regulation.²³ In addition, one former NMR study also indicates that peptide binding into MHC pockets A & B could substantially trigger the dissociation of TAPBPR.¹⁹ Intriguingly, we found that the SL region is structurally heterogeneous due to its highly flexible nature. The SL conformations observed from the MD simulations significantly deviated from the reported crystal structure, with a relative RMSD value reaching up to ~10 Å, which is consistent with the large B-factor of SL observed experimentally.^{3,17,39} Notably, recent NMR studies also suggest a highly flexible SL structure that could even potentially lose contacts with the MHC residues.²²

The jack hairpin region of TAPBPR probably contributes to mediating peptide exchange

In addition to the SL region, one TAPBPR jack hairpin region (residues H231 to R234) is also considered as a critical regulator for peptide exchange, as also suggested by former experimental work.¹⁸ This jack hairpin motif, a highly conserved structural motif in TAPBPR,¹⁷ lies right beneath the floor of the peptide-binding groove and interacts mainly with the β-sheets of the binding groove.^{16,17} As noted above, one stable HB between MHC-I D122 and jack hairpin residue G233 can be formed in both S1 and S2. This HB interaction, however, was abrogated in S3, suggesting that the jack hairpin region might play a role in stabilizing the binding groove, which in turn facilitates the peptide binding. In addition, D122 is also found highly conserved among various MHC-I molecules (see Fig. S11, ESI†). One stable HB between D122 and the β₂m residue W336 can be observed during the whole peptide loading process, consistent with the former conclusion that the W336 is sensitive to the status of peptide loading (see Fig. S9B, ESI†).^{4,40,41} Moreover, one recent deep mutagenesis study also suggests that the substitutions at the G233-site can substantially impair the chaperoning activity of TAPBPR.²² Recent NMR studies also indicate that the binding of TAPBPR or tapasin to MHC-I can profoundly affect the conformation of the peptide-binding groove, likely *via* the jack hairpin region.^{21,22}

Materials and methods

Modelling the partially loaded TAPBPR-bound pMHC-I tertiary complexes

To examine the biological role of TAPBPR in substrate loading into MHC-I, we first constructed three partially loaded pMHC-I in complex with TAPBPR (or TAPBPR-pMHC-I). The TAPBPR conformation was derived from one crystal structure of human TAPBPR bound with one apo mouse H2-D^b (PDB id: 5OPI),¹⁷ and the pMHC-I complexes were obtained from our former computational study.³⁵ Specifically, we previously investigated the complete peptide loading dynamics into MHC-I in the absence of TAPBPR and captured one critical intermediate in

which the peptide N-terminus has docked into the pocket-B of MHC-I but the C-terminus is largely solvent-exposed. Therefore, based on the above identified metastable state, we chose three representative conformations from each of the major free-energy wells (see Fig. S2, ESI†). These pMHC-I structures were then adopted to build the TAPBPR-pMHC-I tertiary complexes. The missing residues of TAPBPR, including L70-Q88, L130-K137, were added using the homology modelling module from the SWISS-MODEL software.^{42–46} Finally, the three TAPBPR-pMHC-I models were subject to energy minimization followed by 10 ns equilibrium MD simulations (see followings for the details of the MD setups).

Obtaining initial peptide loading pathways using steered MD (SMD) simulations

The above energy minimized TAPBPR-pMHC-I complexes were employed as the starting structures for the following SMD simulations,^{47,48} with an aim to derive initial peptide releasing pathways, which presumably, serve as the reverse process of the peptide loading dynamics. In details, for each tertiary complex, we imposed a harmonic potential on the center of mass (com) of the peptide, with varied pulling force constants (200, 230, and 240 kJ mol⁻¹ nm⁻², respectively) and a pulling rate of 0.001 Å ps⁻¹. Three pulling directions were used, defined from the com of three MHC-I β-sheet motifs that serve as the substrate binding floor to the com of the peptide (see Fig. S12, ESI†). Each SMD simulation was conducted using the GROMACS package (version 5.1.2),⁴⁹ with a simulation time of 20 ns. We thus finally collected a total of nine 20 ns SMD trajectories.

Next, based on the above obtained initial peptide loading pathways, we selected 100 representative conformations for the following enhanced MD samplings. To achieve this, we performed geometric clustering for all SMD conformations by firstly projecting the high-dimensional simulation dataset into a low-dimensional phase space using the time-structure independent component (tICA)^{50–52} analysis implemented in the MSMbuilder-3.8 package,⁵³ then clustered the projected dataset into 100 classes by the *k*-centers algorithm. Finally, the center conformations of the obtained 100 classes were used as the starting structures for the subsequent unbiased MD simulations. The input features for tICA were chosen as the following pairwise distances (a total of 906 atom pairs, see Fig. S13A–C, ESI†):

Each substrate-peptide C_α atoms and the C_α atoms of the MHC-I residues E58, Y59, G62, E63, K66, V67, H70, T73, H74, D77, T80, L81, Y84, A139, T143, K146, W147, A150, H151, V152, Q155, L156, Y159, T163, C164, W167, Y171;

The C_α atoms of TAPBPR-SL residues K43-D56 and the C_α atoms of MHC-I residues T73, V76, D77, T80, L81, Y84, A140, T142, T143, K146, W147, A149, A150;

The C_α atoms of TAPBPR-SL residues K43-D56 and the C_α atoms of the peptide residues T384 and V385;

Heavy atoms of pE1 – Non-hydrogen sidechain atoms of Y7, E63, K66, Y159, Y171 of MHC-I;

Heavy atoms of pL2 – Non-hydrogen sidechain atoms of E63, K66 of MHC-I;

Heavy atoms of pA3 – Non-hydrogen sidechain atoms of Y99 of MHC-I.

Shooting extensive unbiased MD simulations

Firstly, starting from each of the above 100 center conformations, we performed one 20 ns MD simulation. All the MD simulations were conducted using the Gromacs-5.1.2 software.⁴⁹ The amber99sb force field was adopted to describe the proteins.⁵⁴ Each TAPBPR-pMHC-I complex was solvated in a triclinic box filled with 35156 TIP3P water molecules. 120 Na⁺ and 98 Cl⁻ ions were added in the aqueous box by replacing random solvent waters to neutralize the system and keep an ionic concentration of 0.15 M. The final system contains a total of 117 336 atoms. For the non-bonded terms, the cut-off values of van der Waals (vdW) and short-range electrostatic interactions were set to 12 Å; the Particle-Mesh Ewald (PME) scheme was employed to calculate the long-range electrostatic interactions;⁵⁵ the LINCS algorithm was used to constrain all the bonds to correct the bond lengths.⁵⁶ Each conformation was firstly subject to energy minimization using the steepest descent method to release the steric clashes. Then, to fully relax the solvent water molecules, one 200 ps restrained NVT MD simulations was performed by constraining all protein heavy atoms. Next, we increased the temperature from 50 to 310 K within 200 ps and equilibrated the whole system in a 20 ns NVT MD simulation at 310 K. The temperature was coupled using the velocity rescaling with a stochastic term.⁵⁷ Notably, to ensure complete coverage of the peptide loading phase space, additional 34 TAPBPR-pMHC-I conformations were selected from the SMD simulations, each was employed for one 20 ns MD simulation. Therefore, we finally achieved a simulation dataset containing 134 20 ns MD simulations.

We further enhanced our samplings by performing another round of MD simulations based on the above simulation dataset. We firstly truncated the first round of simulation datasets by only keeping the last 10 ns conformations for each of the 134 20 ns MD trajectories. We then clustered the remaining dataset into 100 groups using the above-described geometric clustering algorithm, namely, decomposing the high-dimensional space into top-four principal vectors using tICA, followed by the *k*-centers clustering. Finally, each center conformation from the obtained 100 classes was used for two parallel 100 ns MD simulations, with the same MD setups as described above. In addition, we included another 17 100 ns MD simulations to ensure we have sufficient MD samplings to explore all critical conformational space. Thus, we collected a total of 217 100 ns MD simulations with a collective simulation time of 21.7 μs for the final MSM construction.

MSM construction and validation

MSM characterizes the system's dynamics by sampling the system's free energy landscape.⁵⁸ The kinetic model constructs a discrete configuration space based on hundreds of short MD trajectories and yields the transition probability matrix (TPM) \mathbf{T} delineating the transitions between kinetic-relevant states.⁵⁹ Each entry T_{ij} in matrix \mathbf{T} represents the transition probability from state i to state j within a certain lag time τ , which can be

calculated by counting the state-to-state transitions in the raw MD trajectories. Particularly, the lag time τ should be long enough to ensure the complete equilibrium within each state. In this case, the probability of a certain state at time $t + \tau$ is only determined by the state at time t but not the previously visited states, which satisfies Markovian property.⁶⁰ In this way, the long-time dynamics can be acquired by propagating the matrix \mathbf{T} to any future time of interest using the following equation:

$$P(n\tau) = T(\tau)^n P(0) \quad (1)$$

where $P(n\tau)$ and $P(0)$ represent the state population vector at time $n\tau$ and 0, respectively. To ensure the constructed model displays Markovian property, one can plot the implied timescale (τ_k) curves against different lag-time τ values. For each τ value, the corresponding τ_k is defined as:

$$\tau_k = -\tau / \ln \mu_k(\tau) \quad (2)$$

The $\mu_k(\tau)$ represents the k th eigenvalue of the TPM constructed under lag time τ . Each implied timescale curve estimates the transition time between two sets of states and therefore, provides the kinetic information for each transition. If the model is Markovian, the obtained timescale curves tend to level off, from which one can determine the proper lag time τ for the final MSM construction.²⁸

Here, we utilized the “splitting and lumping” strategy to construct MSM.⁶¹ We firstly split the complete simulation dataset based on the geometric properties of all the MD conformations, using the same clustering method as mentioned above. Namely, we employed tICA to decompose the high-dimensional structural features onto the top-four tICs for keeping the slowest conformational dynamics. Then, we further grouped the projected low-dimensional dataset into a varied number of microstates using the *k*-centers clustering method. To aid interpretation of the molecular mechanism underlying the conformational changes, we further lumped the microstates into several macrostates by the PCCA+ method⁶² implemented in MSMbuilder-3.8 package.

To examine how the varied lag time τ and the microstate number might affect the resulting kinetic models,⁶⁰ we constructed several MSMs by applying different lag-time τ (10, 20, and 30 ns, respectively) and the number of microstates (500, 600, 700, and 800, respectively) (see Fig. S14A–C, ESI†). For each pair of parameters, we then plotted the corresponding implied timescale curves. The results suggest that the MSM constructed using a lag-time of 20 ns and microstate number of 500 exhibits a fairly well Markovian property, with the flattened timescale curves converged at ~100 μs. In addition, to further check whether our samplings are sufficient to construct reliable MSM, we truncated each 100 ns MD simulation into two different lengths by keeping the first 80 ns or 90 ns simulation dataset, respectively. For each sub-dataset, we then constructed the corresponding MSM (see Fig. S15A–C, ESI†). The results clearly show that all the calculated timescale curves converge at ~100 μs, indicating that our MD samplings are sufficient to construct a reliable MSM. In addition, we also projected each simulation dataset onto two slowest tICs. The free energy

landscapes show that no additional metastable state appears when the simulation time increases (see Fig. S16A–C, ESI[†]). Moreover, we also calculated the residence probabilities for the 12 most populated microstates according to the constructed MSM and the original MD simulations.⁶³ Our results indicate that the 500-state MSM complies fairly well with the variances of residence probability for these chosen states as observed from the MD simulations, which again, suggests the model is reliable to predict the long-timescale dynamics (see Fig. S17, ESI[†]).

Calculations of the mean first passage time (MFPT) and the stationary populations

The MFPT is defined as the mean time taken for a certain state-to-state transition. We build long-time Monte Carlo (MC) trajectories to calculate the MFPT between each pair of macrostates and the stationary distribution of each macrostate. In details, we generated a 10 ms Monte-Carlo trajectory based on the 500-state TPM constructed under a lag-time of 20 ns, with the corresponding errors estimated using the bootstrapping strategy. That is, we firstly generated 100 new trajectory lists, each containing 217 arbitrary MD simulations selected from the original simulation dataset. Then, for each trajectory list, we constructed a new MSM using the same hyperparameters and calculated the corresponding MFPT and stationary populations. Finally, the mean and errors values of the kinetic and thermodynamic properties were obtained by averaging over the 100 bootstrappings. Additionally, to explore the dominant transition pathways, we adopted the transition path theory (TPT)^{59,64–66} to calculate the net fluxes between all pairs of states according to the 500-state MSM for certain given unbound and bound states. Then, the top transition pathway was determined as the one that exhibits the largest net flux value for the rate-limiting step.

Dynamical cross-correlation matrix (DCCM) analyses

We performed DCCM analysis to examine the correlated dynamics between pocket-B and F in MHC-I in varied contexts. All the analyses were conducted using the bio3d package.^{67,68} We adopted formerly conducted MD simulations for apo MHC-I and pMHC-I. For the TAPBPR-pMHC-I system, we selected all MD trajectories belonging to a given macrostate to calculate the corresponding covariant matrix, and the resulting average matrix was obtained for the final analyses.

Optimal correlation network analyses

We conducted correlation network analysis using the Network-View plugin in visual molecular dynamics (VMD).⁶⁹ We selected the C_α atoms of MHC residues as nodes. Edges connecting each pair of nodes are defined by the inter-atom contacts within 4.5 Å for greater than 75% of the simulation time. Each edge was weighted by correlation values between the two end nodes. Correlations were calculated for the MD trajectory of S1–S3 using the program Carma.^{70,71} Then, the optimal path is identified as the shortest and dynamically correlated path.

Conclusions

In this study, by constructing MSMs based on extensive unbiased MD simulations, we reveal, at the atomic level, the peptide loading dynamics into MHC-I in the presence of TAPBPR. We identified four critical metastable states of peptide involved in its loading process, *i.e.*, S1–S4, among which S1 corresponds to the unbound state, while S2–S4 are N-terminal loaded but C-terminal unloaded states. The dominant loading path was determined as S1 → S2 → S3, in which the second transition limits the overall binding process. Importantly, the peptide binding to pocket-B could significantly promote the SL dissociation from pocket-F, triggered by the allosteric communications between two pockets and direct structural contacts. Meanwhile, the unfolding of the MHC-I α 2-1 region in S3 could also induce altered interactions between α 2-1 helix and TAPBPR, which likely confer to the enhanced couplings between pocket-B and F. Further DCCM analyses uncover the tight structural couplings between pocket-F and B, with TAPBPR bound at pocket-F allosterically reshaping the remote pocket-B in a substrate-receptive conformation.

Author contributions

Honglin Xu, and Lin-Tai Da contributed to the conception of the study, data analyses and manuscript writing; Honglin Xu performed the experiment; Kaiyuan Song helped on the data analyses with constructive discussions.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

L.-T. D. acknowledges the funding support of National Science Foundation of China (grant 22177072). We gratefully acknowledge the computational support from High-Performance Computing of Shanghai Jiao Tong University.

Notes and references

- P. Leone, E. C. Shin, F. Perosa, A. Vacca, F. Dammacco and V. Racanelli, *J. Natl. Cancer Inst.*, 2013, **105**, 1172–1187.
- A. Blees, D. Januliene, T. Hofmann, N. Koller, C. Schmidt, S. Trowitzsch, A. Moeller and R. Tampé, *Nature*, 2017, **551**, 525–528.
- D. H. Margulies, J. Jiang and K. Natarajan, *Curr. Opin. Immunol.*, 2020, **64**, 71–79.
- C. Thomas and R. Tampé, *Curr. Opin. Immunol.*, 2021, **70**, 48–56.
- M. Chen and M. Bouvier, *EMBO J.*, 2007, **26**, 1681–1690.
- G. Dong, P. A. Wearsch, D. R. Peaper, P. Cresswell and K. M. Reinisch, *Immunity*, 2009, **30**, 21–32.

- 7 G. Fleischmann, O. Fisette, C. Thomas, R. Wieneke, F. Tumulka, C. Schneeweiss, S. Springer, L. V. Schäfer and R. Tampé, *J. Immunol.*, 2015, **195**, 4503–4513.
- 8 H. V. Truong and N. G. Sgourakis, *Curr. Opin. Immunol.*, 2021, **70**, 122–128.
- 9 M. S. Teng, R. Stephens, J. A. Lindquist and L. D.-P. A.-J. Trowsdale, *Eur. J. Immunol.*, 2002, **32**, 1059–1068.
- 10 C. Hermann, A. van Hateren, N. Trautwein, A. Neerincx, P. J. Duriez, S. Stevanović, J. Trowsdale, J. E. Deane, T. Elliott and L. H. Boyle, *eLife*, 2015, **4**, e0961.
- 11 L. H. Boyle, C. Hermann, J. M. Boname, K. M. Porter, P. A. Patel, M. L. Burr, L. M. Duncan, M. E. Harbour, D. A. Rhodes, K. Skjødt, P. J. Lehner and J. Trowsdale, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 3465–3470.
- 12 C. Hermann, L. M. Strittmatter, J. E. Deane and L. H. Boyle, *J. Immunol.*, 2013, **191**, 5743–5750.
- 13 G. I. Morozov, H. Zhao, M. G. Mage, L. F. Boyd, J. Jiang, M. A. Dolan, R. Venna, M. A. Norcross, C. P. McMurtrey, W. Hildebrand, P. Schuck, K. Natarajan and D. H. Margulies, *Proc. Natl. Acad. Sci. U. S. A.*, 2016, **113**, E1006–E1015.
- 14 A. Neerincx, C. Hermann, R. Antrobus, A. van Hateren, H. Cao, N. Trautwein, S. Stevanović, T. Elliott, J. E. Deane and L. H. Boyle, *eLife*, 2017, **6**, e23049.
- 15 F. T. Ilca, A. Neerincx, M. R. Wills, M. de la Roche and L. H. Boyle, *Proc. Natl. Acad. Sci. U. S. A.*, 2018, **115**, E9353–E9361.
- 16 J. Jiang, K. Natarajan, L. F. Boyd, G. I. Morozov, M. G. Mage and D. H. Margulies, *Science*, 2017, **358**, 1064–1068.
- 17 C. Thomas and R. Tampé, *Science*, 2017, **358**, 1060–1064.
- 18 F. T. Ilca, A. Neerincx, C. Hermann, A. Marcu, S. Stevanović, J. E. Deane and L. H. Boyle, *eLife*, 2018, **7**, e40126.
- 19 A. C. McShan, K. Natarajan, V. K. Kumirov, D. Flores-Solis, J. Jiang, M. Badstübner, J. S. Toor, C. R. Bagshaw, E. L. Kovrigin, D. H. Margulies and N. G. Sgourakis, *Nat. Chem. Biol.*, 2018, **14**, 811–820.
- 20 F. T. Ilca, L. Z. Drexhage, G. Brewin, S. Peacock and L. H. Boyle, *Cell Rep.*, 2019, **29**, 1621–1632.e1623.
- 21 A. C. McShan, C. A. Devlin, S. A. Overall, J. Park, J. S. Toor, D. Moschidi, D. Flores-Solis, H. Choi, S. Tripathi, E. Procko and N. G. Sgourakis, *Proc. Natl. Acad. Sci. U. S. A.*, 2019, **116**, 25602–25613.
- 22 A. C. McShan, C. A. Devlin, G. I. Morozov, S. A. Overall, D. Moschidi, N. Akella, E. Procko and N. G. Sgourakis, *Nat. Commun.*, 2021, **12**, 3174.
- 23 L. Sagert, F. Hennig, C. Thomas and R. Tampé, *eLife*, 2020, **9**, e55326.
- 24 O. Fisette, G. F. Schröder and L. V. Schäfer, *Proc. Natl. Acad. Sci. U. S. A.*, 2020, **117**, 20597–20606.
- 25 O. Fisette, S. Wingbermühle, R. Tampé and L. V. Schäfer, *Sci. Rep.*, 2016, **6**, 19085.
- 26 R. Anjanappa, M. Garcia-Alai, J. D. Kopicki, J. Lockhauserbäumer, M. Aboelmagd, J. Hinrichs, I. M. Nemtanu, C. Uetrecht, M. Zacharias, S. Springer and R. Meijers, *Nat. Commun.*, 2020, **11**, 1314.
- 27 C. M. Ayres, E. T. Abualrous, A. Bailey, C. Abraham, L. M. Hellman, S. A. Corcelli, F. Noé, T. Elliott and B. M. Baker, *Front. Immunol.*, 2019, **10**, 966.
- 28 G. R. Bowman, X. Huang and V. S. Pande, *Methods*, 2009, **49**, 197–201.
- 29 F. Noé and S. Fischer, *Curr. Opin. Struct. Biol.*, 2008, **18**, 154–162.
- 30 J. D. Chodera and F. Noé, *Curr. Opin. Struct. Biol.*, 2014, **25**, 135–144.
- 31 W. Wang, S. Q. Cao, L. Z. Zhu and X. H. Huang, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2018, **8**, e1343.
- 32 B. E. Husic and V. S. Pande, *J. Am. Chem. Soc.*, 2018, **140**, 2386–2396.
- 33 M. P. Harrigan, M. M. Sultan, C. X. Hernández, B. E. Husic, P. Eastman, C. R. Schwantes, K. A. Beauchamp, R. T. McGibbon and V. S. Pande, *Biophys. J.*, 2017, **112**, 10–15.
- 34 J. R. Abella, D. Antunes, K. Jackson, G. Lizée, C. Clementi and L. E. Kavraki, *Proc. Natl. Acad. Sci. U. S. A.*, 2020, **117**, 30610–30618.
- 35 M. Lin, H. Xu, Y. Shi and L.-T. Da, *CCS Chem.*, 2021, 979–994, DOI: [10.31635/ceschem.021.202000657](https://doi.org/10.31635/ceschem.021.202000657).
- 36 A. Sethi, J. Eargle, A. A. Black and Z. Luthey-Schulten, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 6620–6625.
- 37 Y. Miao, S. E. Nichols, P. M. Gasper, V. T. Metzger and J. A. McCammon, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 10982–10987.
- 38 I. Hafstrand, E. C. Sayitoglu, A. Apavaloei, B. J. Josey, R. Sun, X. Han, S. Pellegrino, D. Ozkazanc, R. Potens, L. Janssen, J. Nilvebrant, P. Nygren, T. Sandalova, S. Springer, A. M. Georgoudaki, A. D. Duru and A. Achour, *Proc. Natl. Acad. Sci. U. S. A.*, 2019, **116**, 5055–5060.
- 39 A. van Hateren and T. Elliott, *Curr. Opin. Immunol.*, 2021, **70**, 138–143.
- 40 C. S. Hee, M. Beerbaum, B. Loll, M. Ballaschk, P. Schmieder, B. Uchanska-Ziegler and A. Ziegler, *Immunogenetics*, 2013, **65**, 157–172.
- 41 M. Beerbaum, M. Ballaschk, N. Erdmann, C. Schnick, A. Diehl, B. Uchanska-Ziegler, A. Ziegler and P. Schmieder, *J. Biomol. NMR*, 2013, **57**, 167–178.
- 42 A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F. T. Heer, T. A.-P. de Beer, C. Rempfer, L. Bordoli, R. Lepore and T. Schwede, *Nucleic Acids Res.*, 2018, **46**, W296–W303.
- 43 S. Bienert, A. Waterhouse, T. A. de Beer, G. Tauriello, G. Studer, L. Bordoli and T. Schwede, *Nucleic Acids Res.*, 2017, **45**, D313–D319.
- 44 N. Guex, M. C. Peitsch and T. Schwede, *Electrophoresis*, 2009, **30**(suppl. 1), S162–S173.
- 45 G. Studer, C. Rempfer, A. M. Waterhouse, R. Gumienny, J. Haas and T. Schwede, *Bioinformatics*, 2020, **36**, 1765–1771.
- 46 M. Bertoni, F. Kiefer, M. Biasini, L. Bordoli and T. Schwede, *Sci. Rep.*, 2017, **7**, 10480.
- 47 B. Isralewitz, M. Gao and K. Schulten, *Curr. Opin. Struct. Biol.*, 2001, **11**, 224–230.
- 48 J. A. Lemkul and D. R. Bevan, *J. Phys. Chem. B*, 2010, **114**, 1652–1660.
- 49 D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. Berendsen, *J. Comput. Chem.*, 2005, **26**, 1701–1718.

- 50 Y. Naritomi and S. Fuchigami, *J. Chem. Phys.*, 2011, **134**, 065101.
- 51 Y. Naritomi and S. Fuchigami, *J. Chem. Phys.*, 2013, **139**, 215102.
- 52 G. Pérez-Hernández, F. Paul, T. Giorgino, G. De Fabritiis and F. Noé, *J. Chem. Phys.*, 2013, **139**, 015102.
- 53 K. A. Beauchamp, G. R. Bowman, T. J. Lane, L. Maibaum, I. S. Haque and V. S. Pande, *J. Chem. Theory Comput.*, 2011, **7**, 3412–3419.
- 54 V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg and C. Simmerling, *Proteins*, 2006, **65**, 712–725.
- 55 U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, L. Hsing and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577–8593.
- 56 B. Hess, H. Bekker, H. J.-C. Berendsen and J. Fraaije, *J. Comput. Chem.*, 1997, **18**, 1463–1472.
- 57 G. Bussi, D. Donadio and M. Parrinello, *J. Chem. Phys.*, 2007, **126**, 014101.
- 58 G. R. Bowman, K. A. Beauchamp, G. Boxer and V. S. Pande, *J. Chem. Phys.*, 2009, **131**, 124101.
- 59 F. Noé, C. Schütte, E. Vanden-Eijnden, L. Reich and T. R. Weikl, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 19011–19016.
- 60 L. Zhang, F. Pardo-Avila, I. C. Unarta, P. P. Cheung, G. Wang, D. Wang and X. Huang, *Acc. Chem. Res.*, 2016, **49**, 687–694.
- 61 J. D. Chodera, N. Singhal, V. S. Pande, K. A. Dill and W. C. Swope, *J. Chem. Phys.*, 2007, **126**, 155101.
- 62 P. Deuflhard and M. Weber, *Linear Algebra Applications*, 2005, **398**, 161–184.
- 63 J. H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J. D. Chodera, C. Schütte and F. Noé, *J. Chem. Phys.*, 2011, **134**, 174105.
- 64 P. Metzner, C. Schutte and E. Vanden-Eijnden, *Multiscale Modeling Simulation*, 2009, **7**, 1192–1219.
- 65 W. L. E and E. Vanden-Eijnden, *J. Statistical Phys.*, 2006, **123**, 503–523.
- 66 A. Berezhkovskii, G. Hummer and A. Szabo, *J. Chem. Phys.*, 2009, **130**, 205102.
- 67 B. J. Grant, A. P. Rodrigues, K. M. ElSawy, J. A. McCammon and L. S. Caves, *Bioinformatics*, 2006, **22**, 2695–2696.
- 68 C. A. Ramirez-Mondragon, M. E. Nguyen, J. Milicaj, B. A. Hassan, F. J. Tucci, R. Muthyalu, J. Gao, E. A. Taylor and Y. Y. Sham, *Int. J. Mol. Sci.*, 2021, **22**, 4619.
- 69 W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graphics*, 1996, **14**(33–38), 27–38.
- 70 P. I. Koukos and N. M. Glykos, *J. Comput. Chem.*, 2013, **34**, 2310–2312.
- 71 N. M. Glykos, *J. Comput. Chem.*, 2006, **27**, 1765–1768.