Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

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

Honglin Xu, Kaiyuan Song and Lin-Tai Da*

Key Laboratory of System Biomedicine (Ministry of Education), Shanghai Center for Systems Biomedicine, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai 200240.

***Corresponding Authors:** darlt@sjtu.edu.cn

This PDF file includes:

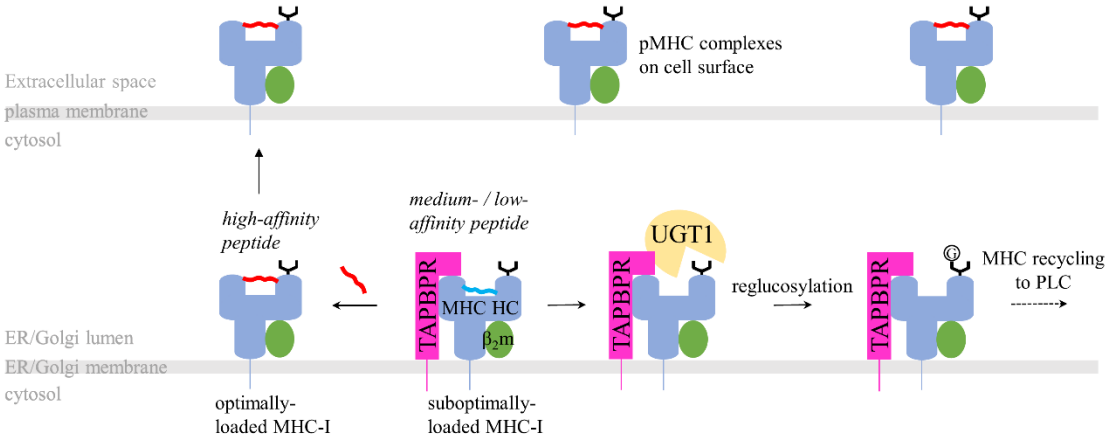


Figure S1. Schematic image illustrating the functional role of TAPBPR in the peptide-presenting pathways.

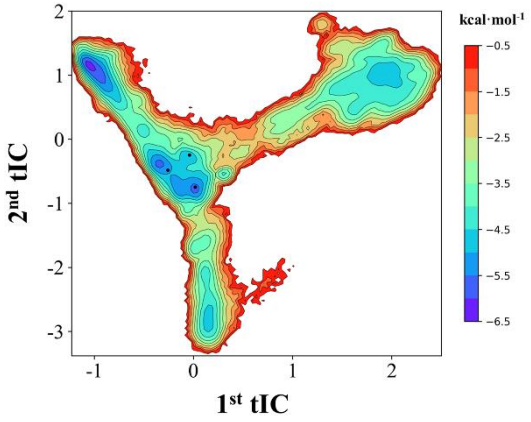


Figure S2. Three representative conformations (highlighted with black dots) were chosen from each of the major free-energy wells for the formerly identified metastable state.[1]

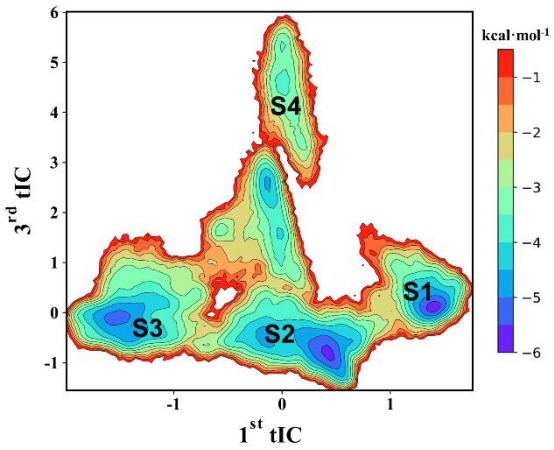


Figure S3. Free energy projection of the MD conformations onto the first tIC and third tIC. The location of each macrostate is labeled.

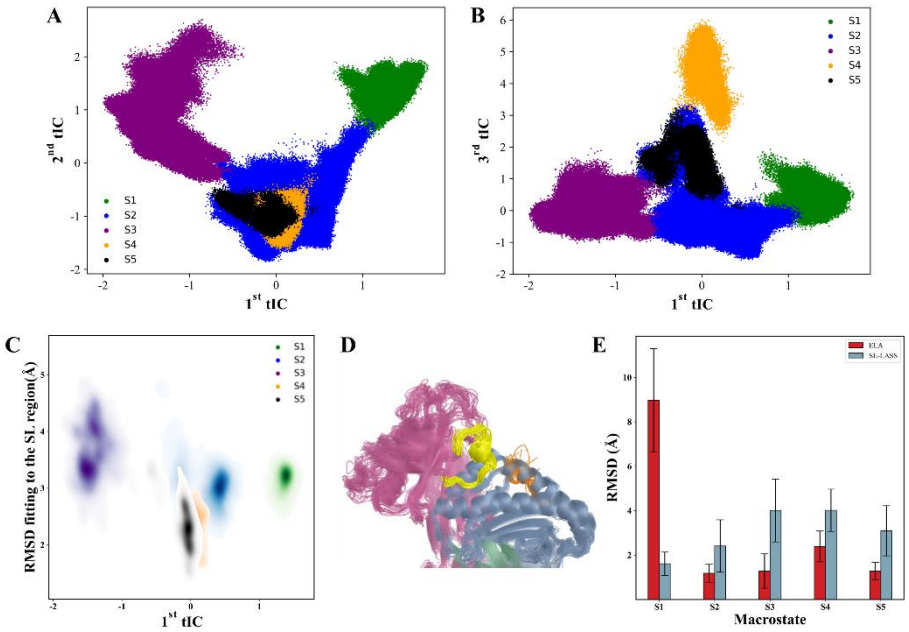


Figure S4. (A-B) Projections of all MD conformations onto tIC1 & tIC2 (A) and tIC1 & tIC3 (B) for the five-state model, colored according to different metastable states. **(C)** Density plot for all MD conformations projected onto two reaction coordinates: the top tIC; the backbone RMSD of the SL residues K43-D56 referenced to the crystal conformation (shown in maintext **Figure 1B**) after fitting to the SL region. **(D)** Representative conformations of S5, with twenty randomly selected structures superimposed. **(E)** Root mean square deviation (RMSD) of the peptide N-terminus (ELA) and the SL-LASS motif for each microstate in the five-state model, calculated for the corresponding C_α atoms with respect to the initial conformation.

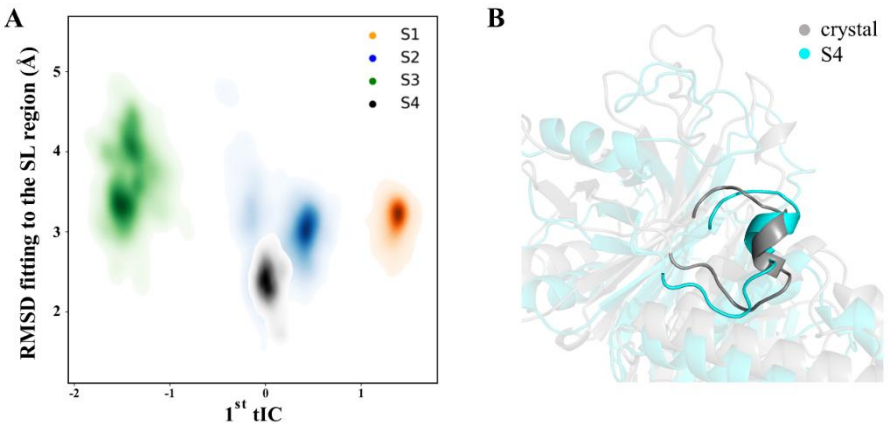


Figure S5. (A) Density plot for all MD conformations projected onto two reaction coordinates: the top tIC; the backbone RMSD of the SL residues K43-D56 referenced to the crystal conformation (shown in Figure 1B) after fitting to the SL region. **(B)** Structural comparison of the SL structure between the crystal form (gray) and one representative S4 conformation (cyan). These two structures are superimposed according to the SL region.

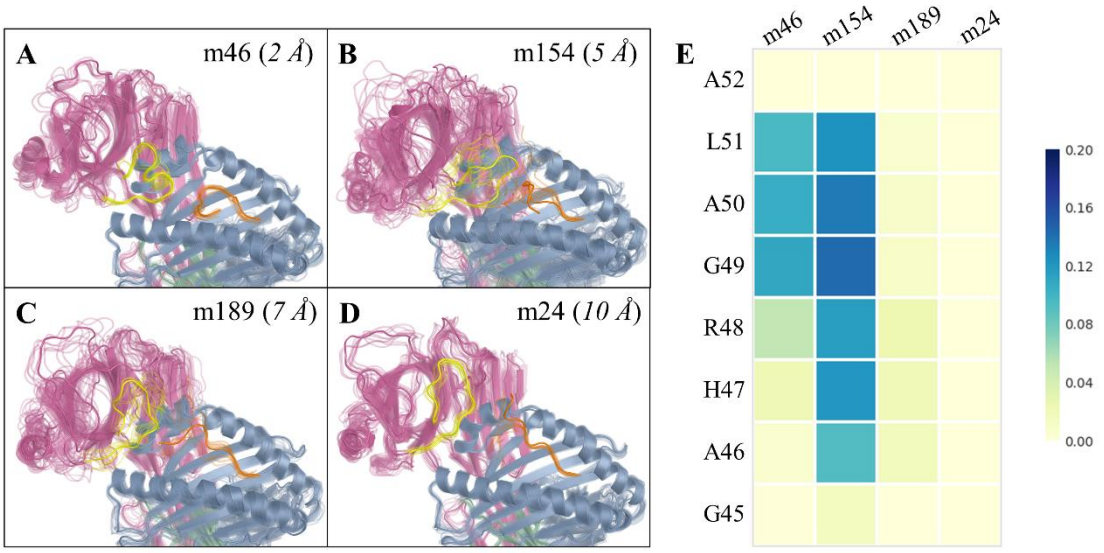


Figure S6. (A)-(D) Representative conformations of each microstate involved in $Path_{S3}$, demonstrating a respective RMSD value of ~2 Å, 5 Å, 7 Å, and 10 Å for the SL residues K43-D56 referenced to the crystal form in maintext **Figure 1B**. For each microstate, ten randomly selected MD conformations were illustrated. **(E)** Helix propensity of the SL K43-D56 motif for each microstate involved in $Path_{S3}$.

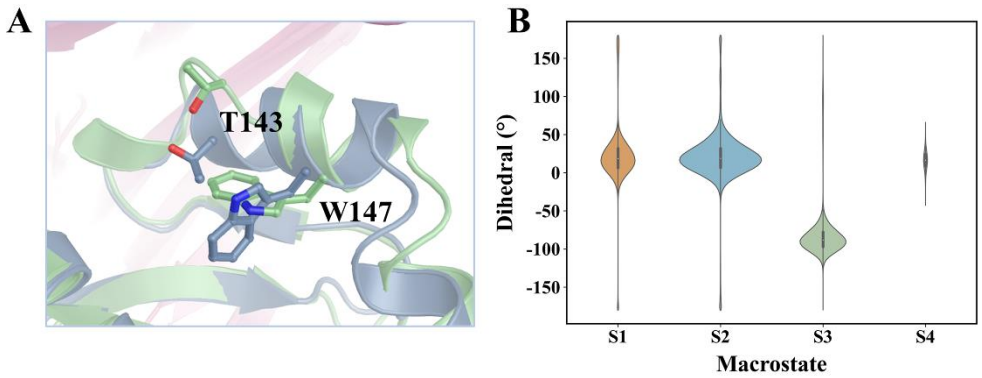


Figure S7. Orientation of the MHC-I W147 sidechain in different states. **(A)** Superposition of two representative conformations from S2 (blue) and S3 (green) where the W147 sidechain adopts distinct orientations. **(B)** The violin plot of the dihedral angle of W147 sidechain in each state.

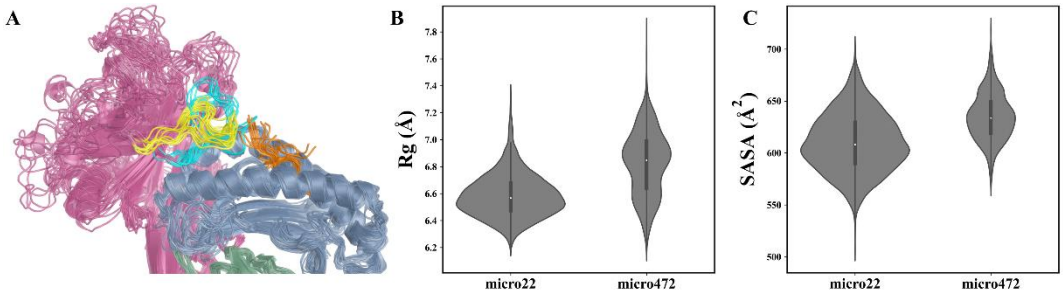
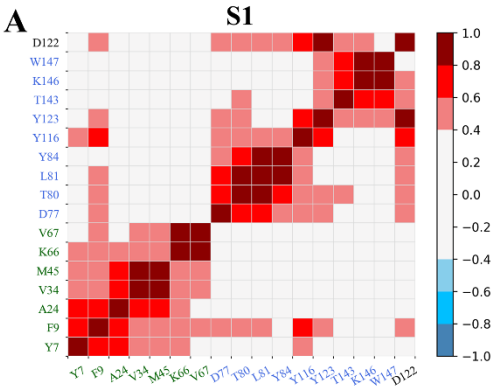


Figure S8. (A) Ten randomly selected representative conformations of micro22 (the SL region in cyan) and micro472 (the SL region in yellow). **(B)** Rg calculations for micro22 and micro472, including SL-S53, MHC residues D77, L81, Y116, Y123, T143, and W147. **(C)** SASA of the SL-LASS motif for micro22 and micro472.

A



B

	D122-G233	D122-W336
S1	0.62	0.97
S2	0.76	0.95
S3	0.06	0.96
S4	0.83	0.62

Figure S9. (A) Cross-correlation analysis for the S1 state, see main text figure 6 for more details. **(B)** The HB occupancy between D122 and G233 & D122 and W336 for each metastable state.

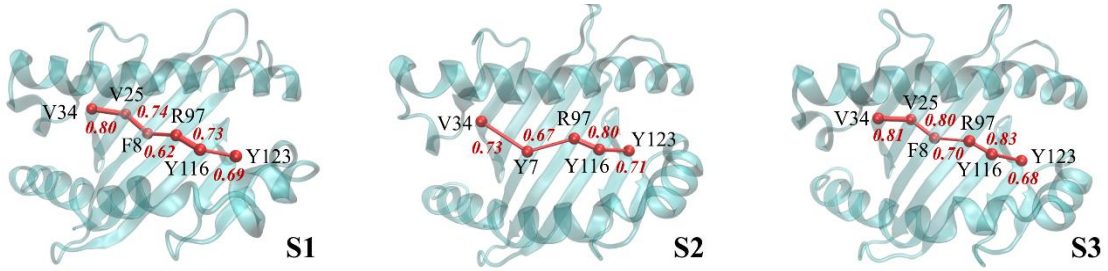


Figure S10. The dominant allosteric pathways between pocket B and F in MHC for S1-S3. The correlations connecting two nodes are indicated with edges, and labeled with the calculated correlation values.

Figure S11. Sequence alignment of several HLA alleles, including HLA-A*02:01 (UniProtKB/Swiss-Prot: P01892), HLA-A*01:01 (UniProtKB/Swiss-Prot: P30443), HLA-B*08:01 (UniProtKB/Swiss-Prot: P30460), HLA-B*15:01 (UniProtKB/Swiss-Prot: P30464), HLA-B*27:02 (UniProtKB/Swiss-Prot: P03989), H2-D^{d} (UniProtKB/Swiss-Prot: P01900), and H2-K^{b} (UniProtKB/Swiss-Prot: P01901). The analyses were conducted using ClustalX.[2, 3] D122 is marked as red star.

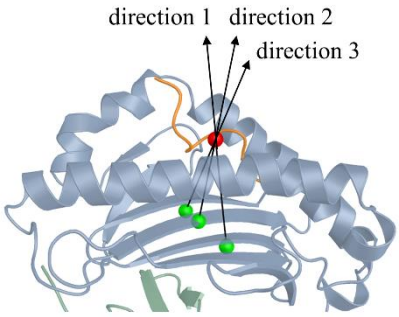


Figure S12. Setup of three pulling directions used for the SMD simulations. Three pulling directions were used, defined from the center of mass (com) of three MHC-I β-sheet motifs (in green spheres) that serve as the substrate binding floor to the com of the peptide (in red sphere). The three reference groups consist of the C_α atoms of MHC-I residues 3-12, residues 94-103, residues 110-118, respectively.

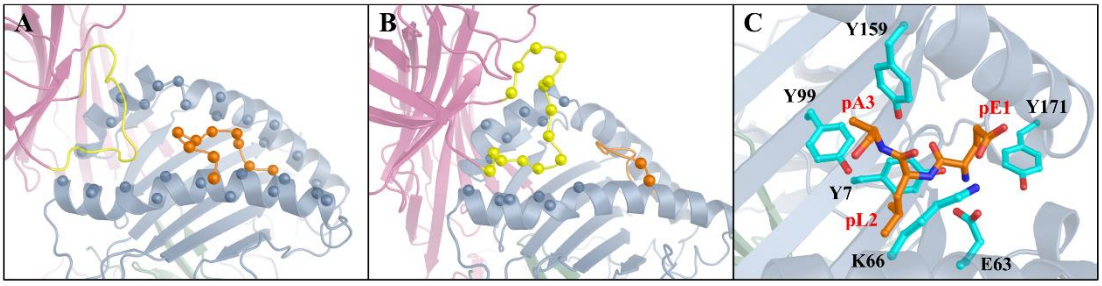


Figure S13. Selected distance pairs used for tICA. **(A)** All the C_α atoms of the peptide (orange spheres) and MHC-I residues (E58, Y59, G62, E63, K66, V67, H70, T73, H74, D77, T80, L81, Y84 on MHC-I α1-helix and A139, T143, K146, W147, A150, H151, V152, Q155, L156, Y159, T163, C164, W167, Y171 on MHC-I α2-helix. blue spheres). **(B)** The C_α atoms of the SL region (residues K43-D56, yellow spheres) and the MHC-I residues (T73, V76, D77, T80, L81, Y84 on MHC-1 α1-helix and A140, T142, T143, K146, W147, A149, A150 on MHC-I α2-helix, blue spheres); The C_α atoms of the SL residues K43-D56 (yellow spheres) and the peptide residues pT9-pV10 (orange spheres). **(C)** Heavy atoms of pE1 and non-hydrogen sidechain atoms of MHC-I residues Y7, E63, K66, Y159, Y171; Heavy atoms of pL2 and non-hydrogen sidechain atoms of MHC-I residues E63, K66 on; Heavy atoms of pA3 and non-hydrogen sidechain atoms of MHC-I residue Y99.

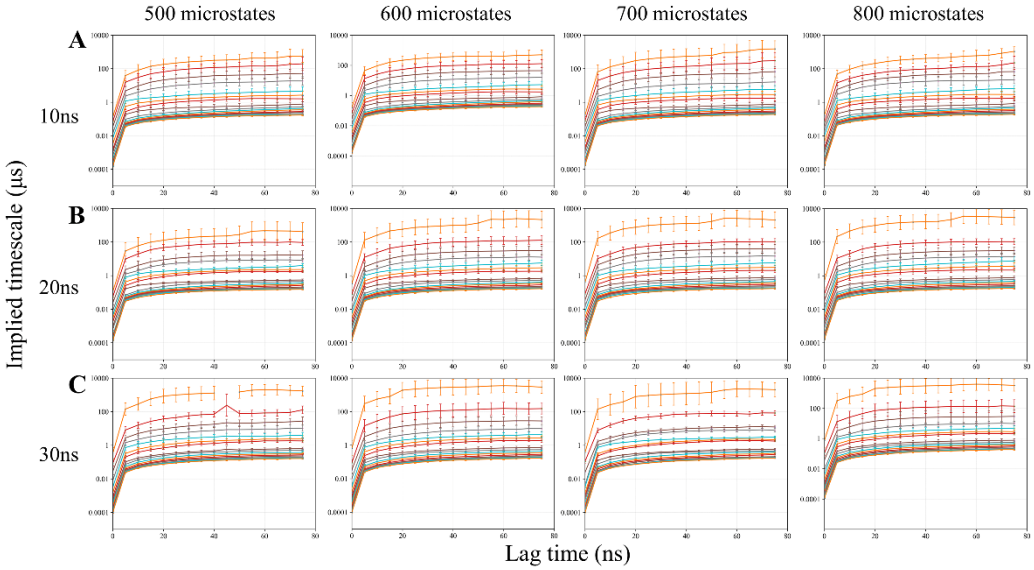


Figure S14. The implied timescale plots at different lag time: 10ns **(A)**, 20ns **(B)**, and 30ns **(C)**. Under each lag time, we constructed MSM by clustering the MD conformations into varied number of microstates, namely, 500, 600, 700, 800.

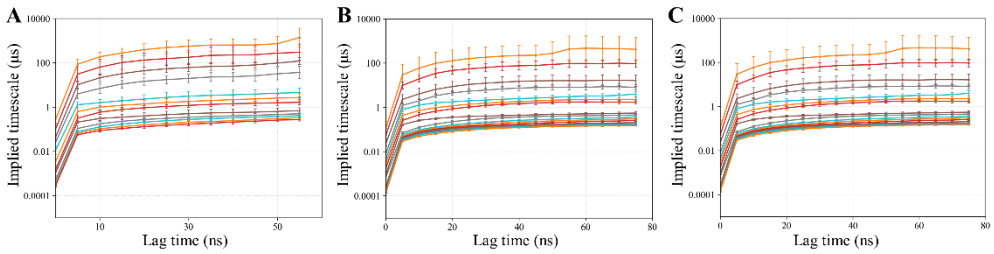


Figure S15. The implied timescale plots for the MSMs constructed using truncated **(A-B)** and complete **(C)** simulation datasets. The former only keeps the first 80-ns (A) or 90-ns (B) simulation dataset for each 100-ns MD simulation.

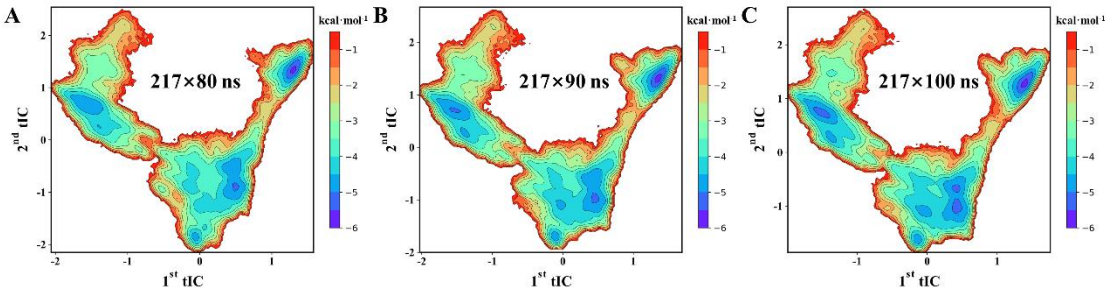


Figure S16. Free energy landscape of different simulation subsets with varied aggregated simulation time:17.4 μs **(A)**, 19.5 μs **(B)**, and 21.7 μs **(C)**, projected onto the same two slowest tICs.

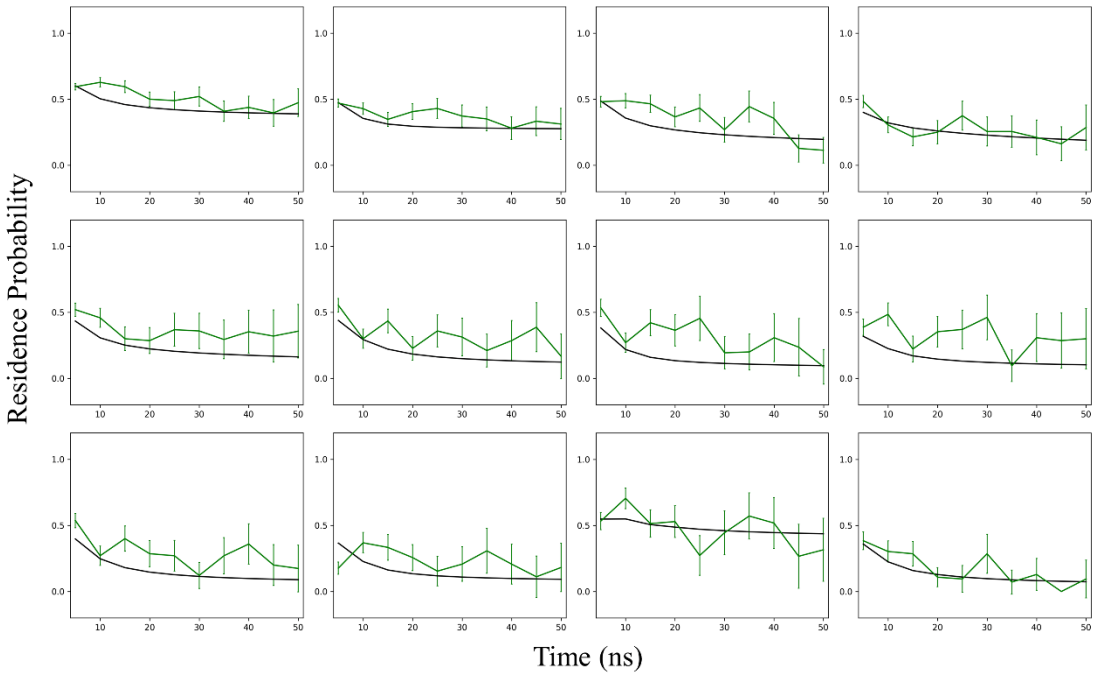


Figure S17. The residence probabilities for the 12 most populated microstates calculated based on the constructed MSM (black) and the original MD simulations (green).

SI References

1. M. Lin, H. Xu, Y. Shi and L.-T. Da, CCS Chem. 2021, 3, 979–994.
2. J. D. Thompson, T. J. Gibson, F. Plewniak, F. Jeanmougin and D. G. Higgins, Nucleic acids research, 1997, 25, 4876-4882.
3. M. A. Larkin, G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson and D. G. Higgins, Bioinformatics (Oxford, England), 2007, 23, 2947-2948.