Supplementary Material

Po-Yu Liang, Xueting Huang, Tibo Duran, Andrew J. Wiemer, Jun Bai

August 12, 2024

1 Molecular Dynamics Simulation Setup

The docking simulation systems for were built using Solution Builder [1] and Multicomponent Assembler [2] from CHARMM-GUI [3], and all systems were simulated with Gromacs 2023 [4]. Visualizations of the simulation results were created using PyMol [5]. The simulation solvent contained TIP3 water molecules with NaCl at a concentration of 0.137 mol/L. The simulation temperature was maintained at 300K to align with the experimental conditions. All simulations were performed following system minimization and equilibration. We performed 5000 steps of minimization using steepest descent algorithm. In the system we used to predict the peptide structure, we performed both NVT and NPT equilibration for 1 ns with 2 fs of time step, using V-rescale [6] for temperature coupling with 0.1 ps of time constant and Parrinello-Rahman [7] for pressure coupling with 2 ps of time constant. As for the docking and umbrella sampling system, we use the default configuration generated by CHARMM-GUI, which includes 5000 steps of minimization using steepest descent algorithm, and NVT equilibration using Nose-Hoover [8] for temperature coupling with 1 ps of time constant. In umbrella sampling, the pulling rate is 9 nm/ns with force constant of 650 $kJ * mol^{-}1 * nm^{-}2$.

2 TIGIT inhibitor sequences

- Positive Group
  - Example Sequence: `CQCSAYFHCMLSVQC`
  - Generated Sequence 1: `PYFHCMLSVQCKTYF`
  - Generated Sequence 2: `PQCSAYFHCMLSVQC*`
  - Generated Sequence 3: `CQCSAYFQCMLSVQC`
- Negative Group
  - Example Sequence: `CNCKRFPQCPLNFLC`
  - Generated Sequence 1: `CNCKRFPQCPQNFLC`
  - Generated Sequence 2: `SNCKRFPQCPLNFLC`
  - Generated Sequence 3: `SSCKRSRQSALSSLS*`

Sequences with an asterisk (*) are selected for further analysis.

Table S1: Average & Standard Deviation of Similarities

Method		ProtT5	ESM-2	BLOSUM	Random
Morgan Fingerprint	3 Sequences	0.8155(0.1416)	**0.8742(0.0809)**	0.5572(0.1035)	0.3918(0.1130)
	5 Sequences	0.7982(0.1535)	**0.8745(0.0810)**	0.5483(0.1030)	0.3915(0.1127)
	10 Sequences	0.7697(0.1688)	**0.8752(0.0819)**	0.5359(0.1026)	0.3921(0.1128)
	length<10	0.7228(0.1760)	**0.8449(0.0851)**	0.5393(0.1056)	0.3518(0.1045)
	10<length<15	0.7890(0.1575)	**0.8889(0.0705)**	0.5244(0.0995)	0.4000(0.1047)
	15<length	0.8350(0.1395)	**0.9134(0.0658)**	0.5408(0.1007)	0.4569(0.1062)
RDKit Descriptor	3 Sequences	**0.9915(0.0232)**	0.9497(0.0974)	0.9297(0.0436)	0.8874(0.0557)
	5 Sequences	**0.9894(0.0277)**	0.9518(0.0953)	0.9281(0.0443)	0.8871(0.0556)
	10 Sequences	**0.9861(0.0326)**	0.9550(0.0922)	0.9254(0.0447)	0.8872(0.0555)
	length<10	**0.9837(0.0370)**	0.9456(0.1071)	0.9255(0.0504)	0.8761(0.0605)
	10<length<15	**0.9869(0.0305)**	0.9571(0.0869)	0.9193(0.0429)	0.8841(0.0517)
	15<length	**0.9899(0.0244)**	0.9704(0.0600)	0.9320(0.0335)	0.9105(0.0421)
Sequence QSAR	3 Sequences	0.9926(0.0157)	**0.9961(0.0035)**	0.9804(0.0102)	0.9603(0.0273)
	5 Sequences	0.9911(0.0174)	**0.9961(0.0035)**	0.9797(0.0106)	0.9602(0.0273)
	10 Sequences	0.9888(0.0198)	**0.9960(0.0035)**	0.9787(0.0113)	0.9602(0.0270)
	length<10	0.9837(0.0244)	**0.9939(0.0040)**	0.9766(0.0113)	0.9516(0.0301)
	10<length<15	0.9915(0.0152)	**0.9973(0.0014)**	0.9792(0.0111)	0.9639(0.0230)
	15<length	0.9951(0.0110)	**0.9986(0.0008)**	0.9816(0.0112)	0.9713(0.0204)

**Values are shown with four digits to highlight QSAR similarity differences.*
***Values are presented in the format "mean (std)"*

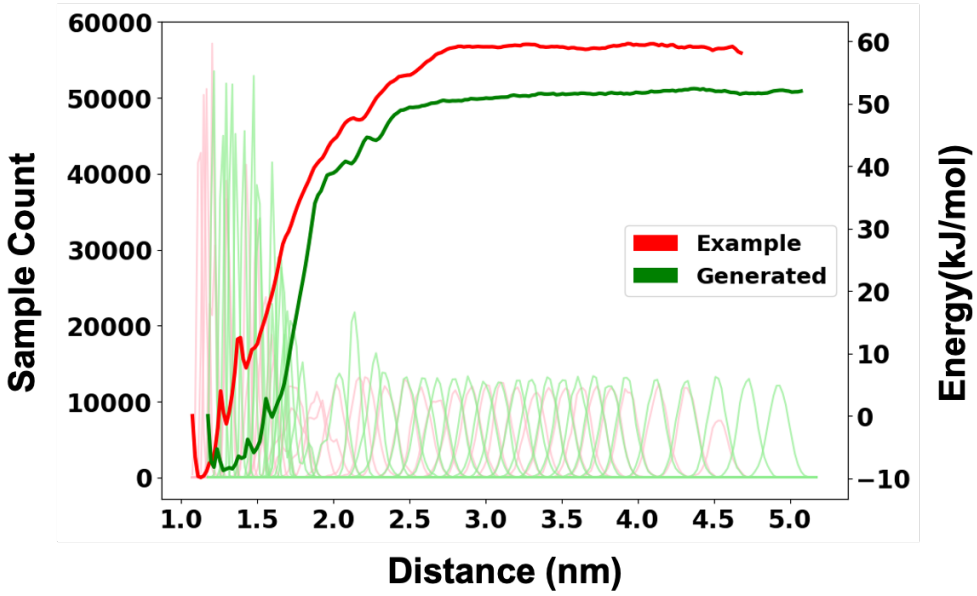

Figure S1: Umbrella Sampling Result of Positive Example

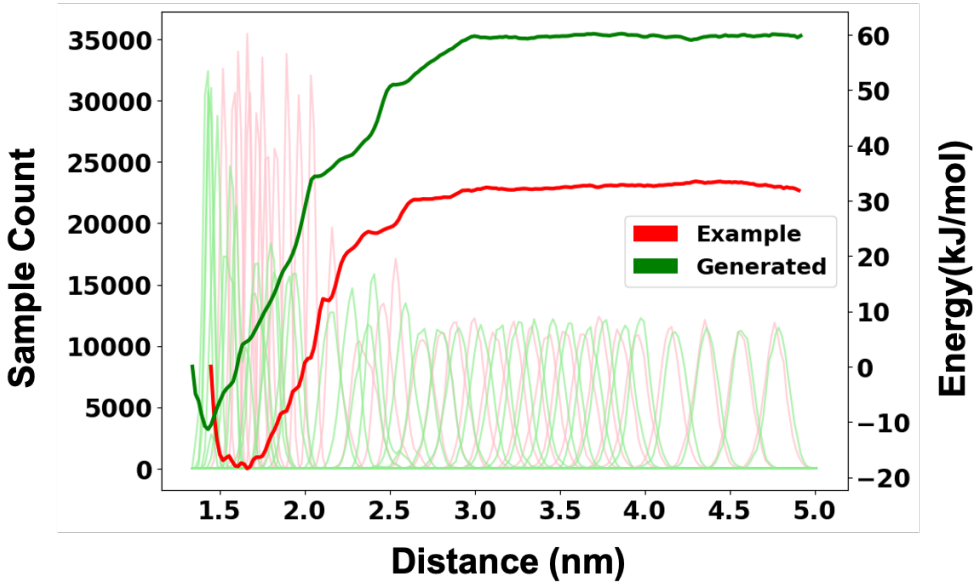

Figure S2: Umbrella Sampling Result of Negative Example

Figure S3: Residue Distances of Positive Example

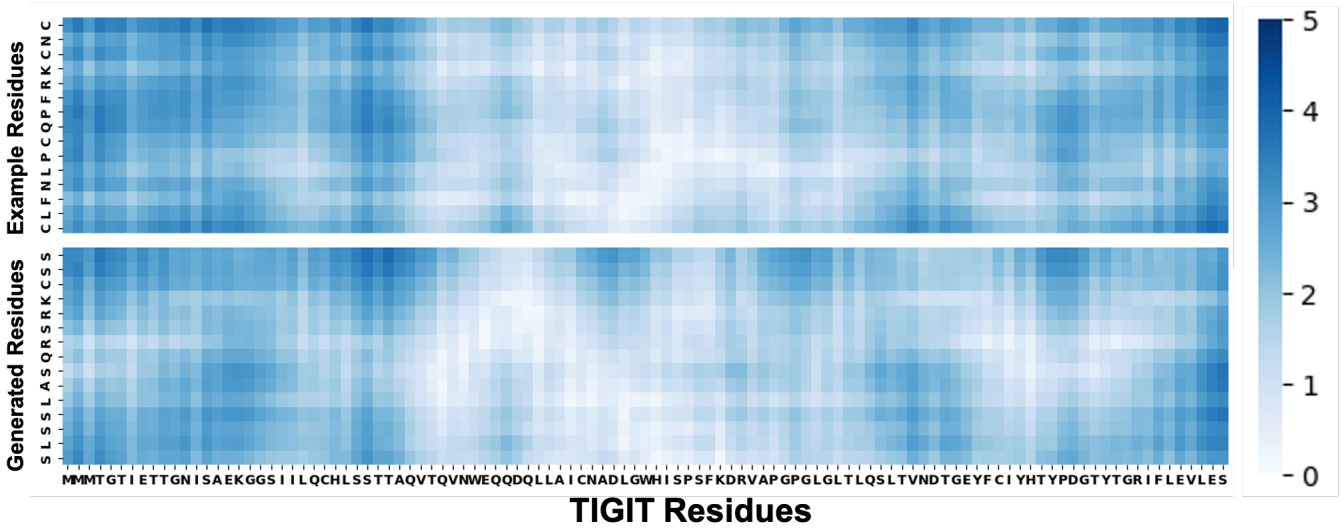

Figure S4: Residue Distances of Negative Example

References

[1] J. Lee, X. Cheng, S. Jo, A. D. MacKerell, J. B. Klauda, and W. Im, "Charmm-gui input generator for namd, gromacs, amber, openmm, and charmm/openmm simulations using the charmm36 additive force field," *Biophysical journal*, vol. 110, no. 3, p. 641a, 2016.

[2] N. R. Kern, J. Lee, Y. K. Choi, and W. Im, "Charmm-gui multicomponent assembler for modeling and simulation of complex multicomponent systems," *Nature Communications*, vol. 15, no. 1, pp. 1–14, 2024.

[3] S. Jo, T. Kim, V. G. Iyer, and W. Im, "Charmm-gui: a web-based graphical user interface for charmm," *Journal of computational chemistry*, vol. 29, no. 11, pp. 1859–1865, 2008.

[4] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl, "Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers," *SoftwareX*, vol. 1, pp. 19–25, 2015.

[5] PyMOL, "Pymol: Molecular visualization system," 2024. Accessed: 2024-07-22.

[6] G. Bussi, D. Donadio, and M. Parrinello, "Canonical sampling through velocity rescaling," *The Journal of chemical physics*, vol. 126, no. 1, 2007.

[7] M. Parrinello and A. Rahman, "Polymorphic transitions in single crystals: A new molecular dynamics method," *Journal of Applied physics*, vol. 52, no. 12, pp. 7182–7190, 1981.

[8] P. H. Hünenberger, "Thermostat algorithms for molecular dynamics simulations," *Advanced computer simulation: Approaches for soft matter sciences I*, pp. 105–149, 2005.